Estimating multivariate density function of mixed measurement error data

by

Linruo Guo

B.S., Help University, Malaysia, 2015

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

Density estimation has been a long frontline research area in nonparametric smoothing. However, real applications oftentimes see the data contaminated with different types of measurement errors. Further data analysis, therefore, should take care of these errors to have a reliable statistical inference procedure. In this proposal, nonparametric density estimation for the data contaminated super-smooth, ordinary-smooth, Berkson measurement errors will be thoroughly investigated. Classical kernel and deconvolution kernel smoothing are used as building blocks to construct the estimators.

In the first part, we propose a nonparametric mixed kernel estimator for a multivariate density function and its derivatives when the data are contaminated with different sources of measurement errors. The proposed estimator is a mixture of the classical and the deconvolution kernels, accounting for the error-free and error- prone variables, respectively. Large sample properties of the proposed nonparametric estimator, including the order of the mean squares error, the consistency, and the asymptotic normality, are discussed. The optimal convergence rates among all nonparametric estimators for different measurement error structures are derived, and it is shown that the proposed mixed kernel estimators achieve the optimal convergence rate. A simulation study is conducted to evaluate the finite sample performance of the proposed estimators.

In the second part, we consider the nonparametric estimation for the joint density function of two random variables, when one variable is contaminated with Berkson measurement error, and another variable can be observed directly. Two estimators are proposed with or without applying the kernel smoothing for the data with Berkson measurement error. Mean squared errors are calculated for both estimators. Large sample properties, including weak consistencies, strong consistencies, uniform strong consistencies in probability, and asymptotic normality are derived. In addition, we develop a method for bandwidth selection in the

kernel estimate of the probability density using the least squares cross-validation method. The performance of this method is further assessed by a simulation study.

Estimating multivariate density function of mixed measurement error data

by

Linruo Guo

B.S., Help University, Malaysia, 2015

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Dr. Weixing Song

Copyright

Abstract

Density estimation has been a long frontline research area in nonparametric smoothing. However, real applications oftentimes see the data contaminated with different types of measurement errors. Further data analysis, therefore, should take care of these errors to have a reliable statistical inference procedure. In this proposal, nonparametric density estimation for the data contaminated super-smooth, ordinary-smooth, Berkson measurement errors will be thoroughly investigated. Classical kernel and deconvolution kernel smoothing are used as building blocks to construct the estimators.

In the first part, we propose a nonparametric mixed kernel estimator for a multivariate density function and its derivatives when the data are contaminated with different sources of measurement errors. The proposed estimator is a mixture of the classical and the deconvolution kernels, accounting for the error-free and error- prone variables, respectively. Large sample properties of the proposed nonparametric estimator, including the order of the mean squares error, the consistency, and the asymptotic normality, are discussed. The optimal convergence rates among all nonparametric estimators for different measurement error structures are derived, and it is shown that the proposed mixed kernel estimators achieve the optimal convergence rate. A simulation study is conducted to evaluate the finite sample performance of the proposed estimators.

In the second part, we consider the nonparametric estimation for the joint density function of two random variables, when one variable is contaminated with Berkson measurement error, and another variable can be observed directly. Two estimators are proposed with or without applying the kernel smoothing for the data with Berkson measurement error. Mean squared errors are calculated for both estimators. Large sample properties, including weak consistencies, strong consistencies, uniform strong consistencies in probability, and asymptotic normality are derived. In addition, we develop a method for bandwidth selection in the

kernel estimate of the probability density using the least squares cross-validation method. The performance of this method is further assessed by a simulation study.

Table of Contents

List of Figures

Acknowledgments

Throughout my doctoral studies, I came across a number of obstacles that were difficult to overcome on my own. Many wonderful people showed up and offered me a great deal of assistance and support, which I greatly appreciate. I'd like to thank them from the bottom of my heart.

First and foremost, I'd like to express my gratitude to Professor Weixing Song, whose immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I am extremely thankful to his invaluable advice, continuous support and patience during my PhD study.

In addition, I would like to thank Dr. Wei-Wen Hsu, Dr. Juan Du, and Dr. Bala Natarajan, members of my dissertation committee, for their time and effort, as well as their insightful comments during my defense.

I would also like to thank the Department of Statistics at Kansas State University for supporting me as a GTA from 2017 Fall to 2022 Spring. It was an extremely beneficial experience in which I not only learned about statistics but also how to teach them. As a result, I was able to obtain my first job, which I really appreciate.

Further more, I am grateful for the opportunity to work with such talented fellows. At K-state, we have created a fulfilling and happy doctoral life and formed a sympathetic friendship with each other. I sincerely thank Weijia Jia, Shiqiang Jin, Jia Liang, Chixiang Chen, Huaiyu Zhang, Jianfeng Chen, shengnan Chen and Weikang Duan for their guidance and help.

Additionally, I would like to express my sincere gratitude to all of the professors who have diligently taught us over the past six years. It is because of their rigorous training that I have been able to grow and improve. They've broadened my horizons and provided a solid foundation for my dissertation and professional development.

Last but not the least, I would like to express my appreciation to my families. They have always shown me the greatest selflessness in their love and care, and it is their life and spirit support that has given me the courage to face one challenge after another. Many thanks to my parents, father-in-law, mother-in-law, and especially my wife, Ms. Chen Zhao, whose understanding, tolerance, and respect are important spiritual pillars for me; and finally, I thank God for my lovely sons, Michael and Max, they are the source of strength for me to always be optimistic and progressive in my life!

Chapter 1

Introduction

An estimation of a probability density function is composed of estimating an unobservable underlying probability function from observed data. It is generally thought of that the data represent a random sample from a population, while an unseen density function represents the distribution of a large population according to its density. Also, knowing the probability density function of a statistical population can be useful in determining if an observation is likely, or so unlikely that it should be considered an anomaly or outlier. Additionally, it allows you to select the right learning method if you require that the input data have a certain probability distribution. Different methods are used to estimate density, including Parzen windows, a variety of data clustering techniques, and vector quantization. A re-scaled histogram is the most basic density estimation method.

Suppose that $X_1, \cdots, X_n$ is a set of continuous random variable having common density f. The parametric approach to estimate f involves assuming that f belongs to a parametric family of distributions, such as normal or gamma family, and then estimating the unknown parameters using, for example, the maximum likelihood estimation procedure. On the other hand, a nonparametric density estimator assumes no pre-specified functional form for f. Especially for those data sample that may not resemble a common probability distribution or cannot be easily made to fit the distribution.

The most common nonparametric approach for estimating the probability density function of a continuous random variable is the kernel smoothing, or kernel density estimation. It was first introduced in the literature for univariate data in the Rosenblatt (1956) and Parzen (1962) and can be viewed as a generalization of histogram density estimation with better statistical properties. Multivariate kernel density estimation has reached comparable maturity levels with its univariate counterparts after years of research in the 1990s and 2000s.

Kernel smoothing is a mathematical function that returns a probability for a given value of a random variable. In essence, the kernel smooths or interpolates the probabilities across the range of outcomes for a random variable, such that the sum of probabilities equals one, this is a precondition for well-behaved probabilities. Suppose that we want to visualize a set of data graphically. Then kernel density estimation is an attractive procedure to achieving this goal since kernel estimators are simple and mathematically tractable, allowing one to gain a deep understanding of their properties without knowing highly sophisticated mathematics. In addition, kernel smoothing is an effective method for solving a variety of important problems with simplicity and reliability. It is a well known data analytic tool which provides a very effective way of showing structure in a set of data at the beginning of its analysis. For example, when the data has two peaks (bimodal distribution) or many peaks (multimodal distribution). General parametric estimation is not feasible and it is impossible to capture the significance by imposing a unimodal parametric mode. Since one of the main goals of data analysis is to highlight important structures in the data, it is desirable to have a density estimator, such as kernel smoothing, that does not assume that the density has a particular function form. In this case, the distribution will still have parameters but are not directly controllable in the same way as regular probability distributions. It estimates the density using all observations in the random sample and treating all observations in the sample as "parameters".

The univariate kernel density estimator is defined as follow

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

Here K is the kernel function which satisfying $\int K(x)dx = 1$, and h is a sequence of positive numbers tending to 0 as the sample size goes to infinity, which we call the bandwidth or window width. One can think of the kernel estimate is constructed by centering a scaled kernel at each observation. The estimated value at point x is simply the average of the n kernel ordinates at that point. By combining contributions from each data point, in regions where there are many observations, and it is expected that the true density is relatively large, the kernel estimate should also have a relatively large value. When there are relatively few observations in the region, the opposite is expected to occur.

The parameter h also known as smoothing parameter, it controls the number of samples or window of samples used to estimate the probability for a given sample. The choice of value for the bandwidth is critical. A large bandwidth may produce a coarse density estimate with little details, which is said to be oversmoothed. For a small window width, the detail may be too fine and not smooth enough or general enough to correctly cover new or unseen samples. As a result, the averaging process performed at each point is based on relative few observations and the estimate of f will be very rough. This estimate pays too much attention to the particular data set and limits the variation across samples, which is known as undersmoothed.

Note that the choice of the kernel function is not particularly important. The contribution of samples within the window can be shaped using different kernel functions. Usually K is chosen to be a unimodal probability density function that is symmetric about zero, e.g. normal, biweight, etc., with different effects on the smoothness of the resulting density function. This guaranties that $\hat{f}(x; h)$ itself is also a density.

Apparently, in order to make accurate statistical inferences, the data should be collected as precisely as possible. However, in real applications, there are often mismeasured values in data sets or even values that cannot be directly observed. As a result, we observe contaminated covariates, which could be due to random errors or systematic errors.

Especially in regression where some regressors are measured with errors, estimation based on the standard assumption yields inconsistent results, showing that parameter estimations do not converge to true values even in large samples. Attenuation bias occurs when simple linear regression results in an underestimating of the coefficient. In non-linear models, determining the bias direction is likely to be more complex. More technical details can be found in Fuller (2009). Therefore, we need to consider the regression model that account for measurement errors in the independent variables.

However, much of the literature on measurement error is based on what is known as classical measurement error, in which the truth is measured with additive error, typically with constant variance. Under the classical error framework, we often denote X to be the true variable and Z to be the observed contaminated variable. Then the classical measurement error model states that $Z = X + U$. In this model, X and U is assumed to be independent and U has mean 0. Therefore, Z is an unbiased measure of X. The error structure of U could be homoscedastic (constant variance) or heteroscedastic.

The examples of the classical measurement error model are abundant. For example, in the National Cancer Institute's OPEN study Subar et al. (2001), the research interest is in measuring the logarithm of dietary protein intake, denoted by X. However, the true long-term log-intake X cannot be observed in practice. Instead, the researchers measured a biomarker of log-protein intake, namely urinary nitrogen, denoted by Z. Based on feeding studies, it is evident that the protein biomarker accurately reflects actual protein intake with added variability.

What we see in the classical measurement error model is that the observed variable equals the true variable plus a measurement error. This leads to a greater variability in observed variable than in true variable. In contrast, another different measurement error structure also appears in the literature. In this structure, the true variable is assumed to be equal to the observed variable plus the measurement error, that is, $X = Z + U$, where Z and U are independent, and $E(U) = 0$. This type of error is called the Berkson measurement error. In this case, true variable is more variable than the observed variable.

An example of Berkson error can be found in Young et al. (1975). In Young et al. (1975), 58 subjects were randomly assigned to various nitrogen intake levels. Subjects are instructed to consume only prepared meals with known nitrogen contents in order to control their intake. It is impossible to observe the actual intake, X, due to variations in nutritional content, serving size, etc. The target intake is a fixed value, so this is a case of Berkson error. There are additional practice examples of measuring errors in Carroll et al. (1995).

The "Triple Whammy", mentioned by Carroll and Hall (1988), describes three main effects of measurement error on statistical analysis. It may not only result in biased parameter estimation but also power losses in testing, thus adversely affecting the ability to detect clinically meaningful relationships between variables. In the case of nonlinear models, features of the data are masked, as a result, the graphical representations of those data do not truly represent how the explanatory variable influences the response variable.

In measurement error modeling, there has been growing interest in estimating the density function f of X from the contaminated data Z. For example, if X represents the usual intake of a particular nutrient, then knowing its density can help us to better understand the dietary patterns of individuals within a population. In this context, the deconvolution kernel density estimator from Stefanski and Carroll (1990) and Carroll and Hall (1988) is the most popular nonparametric estimator of f. Under the classical measurement error frame work, the deconvolution kernel density estimator is constructed by exploiting the independence between X and U. The independence implies that $\phi_Z(t) = \phi_X(t)\phi_U(t)$, where for a generic random variable Y, $\phi_Y(t) = E(e^{iYt})$. Also, for an absolutely integrable function g, its Fourier transform has the form of $\int e^{itx}g(x)dx$, where i represents the complex number such that $i^2 = -1$. Thus, if we further assume that $\phi_U(t) \neq 0$ for all $t \in \mathbb{R}$, then $\phi_X(t) = \phi_Z(t)/\phi_U(t)$. By the Fourier inversion theorem, we conclude that

$$f_X(x) = \frac{1}{2\pi}\int e^{-itx}\phi_X(t)dt = \frac{1}{2\pi}\int e^{-itx}\frac{\phi_Z(t)}{\phi_U(t)}dt.$$

Because f_U is known, ϕ_U is known. Leaving only ϕ_Z is unknown, but it can be easily estimated using the observations on Z, such as applying the empirical characteristic function $\hat{\phi}_Z(t) = n^{-1}\sum_{j=1}^{n} e^{itZ_j}$. Then, by plugging $\hat{\phi}_Z(t)$ back, the estimator for $f_X(x)$ can be constructed. However, for large $|t|$, $\phi_U(t) \to 0$ as $|t| \to \infty$, it causes $\hat{\phi}_Z(t)/\phi_U(t)$ not integrable. As a way to overcome unreliable tail fluctuations, Stefanski and Carroll (1990) and Carroll and Hall (1988) introduced a weight function $w(t)$ such that it is close to one when $\hat{\phi}_Z(t)$ is reliable, and close to zero otherwise. They define

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\hat{\phi}_Z(t) w(t)}{\phi_U(t)} dt,$$

where $w(t) = \phi_K(ht)$, with a bandwidth parameter h and a univariate kernel function K. Subsequently, Stefanski and Carroll (1990) and Carroll and Hall (1988) defined the deconvolution kernel density estimator as follow:

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\hat{\phi}_Z(t) w(t)}{\phi_U(t)} dt = \frac{1}{nh} \sum_{j=1}^{n} K_n \left(\frac{x - Z_j}{h} \right),$$

where the deconvolution kernel K_n is defined by

$$K_n(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_K(t)}{\phi_U(t/h)} dt.$$

with following conditions imposed to guarantee desirable statistical properties of the resulting estimator,

(1) $\phi_U(t) \neq 0$ for all t;

(2) $\int |\phi_X(t)| dt < \infty$;

(3) $\sup_{t \in \mathbb{R}} |\phi_K(t)/\phi_U(t/h)| < \infty$ and $\int |\phi_K(t)/\phi_U(t/h)| dt < \infty$.

In real applications, both error-prone and error-free variables are often included in the model. The discussion on the estimation of the joint density function of both variables or error-prone and error-free is limited in the literature. As far as we are aware, there has not been any discussion of the true multivariate density estimation in the literature, when all types of measurement errors shown in the data, and the data are multidimensional.

In this dissertation, we will try extend the existing methods to handle the multidimensional cases with various type of measurement errors. In chapter 2, we aim to find out which type of measurement errors and which variable within each measurement error type, play important role in the determination of the optimal convergence rate of nonparametric estimators of joint density or its derivatives. We will show that the mixed kernel estimator proposed in this dissertation achieves the optimal convergence rate. The mixed kernel density estimator can be readily used to construct estimate procedures for the regression functions when the predictors in the regression model are contaminated different types of measurement errors.

In chapter 3, we consider the nonparametric estimation for joint density function of two univariate random variables, when one variable is contaminated with the Berkson measurement error and another variable are error-free. Two estimators with different smoothing schemes are proposed.

The large sample features of all the estimators suggested in this dissertation, including consistencies and asymptotic normality, are thoroughly examined. In addition, simulation studies are also included to further prove the proposed results.

Chapter 2

Estimating Multivariate Density and Its Derivatives for Mixed Measurement Error Data

2.1 Introduction

As one of the most commonly used nonparametric smoothing techniques, the kernel-based local smoothing estimation procedure has been widely used in modeling the relationship between a scalar response variable and a set of covariates. The building block for the kernel-based estimation procedure is the kernel estimator of the density function of the covariates. Extensive introductions on these topics can be founded in Wand and Jones (1995), Fan and Gijbels (1996), Wang (2011), among others. In real applications, the random system of interest might involve many variables, some variables can be observed directly, and some variables cannot due to the instrumental flaws or human errors. For those variables that cannot be observed, we may observe their surrogates with different measurement error structures. In the univariate case when the variable is measured with measurement errors, Stefanski and Carroll (1990) proposed a deconvolution kernel estimator to estimate the density function of the true latent variable. Since then, more extensive research on this subject has been conducted, for example, optimal convergence rates for the deconvolution kernel density estimators are discussed in Carroll and Hall (1988), Fan (1991a), the asymptotic normality of the derivative of the deconvolution kernel density estimator is proved in Fan (1991b). Later, the deconvolution technique is extended to multivariate setup, see Fan and Masry (1992), Masry (1991), Masry (1993a), Masry (1993b), Masry (1993c) and Tang (1994).

However, real applications often see that both the error-free and error-prone variables are shown in the model, and the discussion on the estimation of the joint density function seems scant in literature. Recently, Shi et al. (2020) studied both nonparametric density estimators and regression function estimators in a statistical model when some predictors are free of measurement errors and some predictors are contaminated by the classical measurement errors with different smoothness. Note that Shi et al. (2020) only discussed a relative simple scenario in which only two one-dimensional predictors are presented in the model, one of these two predictors has no measurement error, and another has either the ordinary or the super-smooth measurement error. It is shown that in both cases, the optimal convergence rates of the proposed mixed deconvolution kernel estimators are completely determined by the smoothness of the measurement errors. To our best knowledge, the literature has no discussion on the true multivariate cases, in particular, when all types of measurement error structures are shown in the data and the predictors with each type of measurement error are multidimensional. In this paper, we will try to fill this void. The contribution made in this paper include

(1). Identifying which measurement error type, and which variable within each measurement error type, plays dominated role in determining the optimal convergence rate of nonparametric estimators of the joint density or its derivatives;

(2). Proposing a mixed kernel estimator and to check if it achieves the optimal convergence rate, and

(3). Laying a solid foundation for estimating the regression functions when all types of measurement errors appear in the regression model with multiple predictors.

Researches in statistical models with certain other types of measurement error structures can be found in literature, for examples, Delaigle (2007) proposed a nonparametric density estimator for the data with a mixture of Berkson and classical errors, while Carroll et al. (2007a) further considered a nonparametric regression estimation procedures in a nonparametric regression model when the covariate is contaminated by a mixture of Berkson and classical errors.

The paper is organized as follow. Notations for multivariate indices, and two types of measurement errors, the ordinary smooth and the super-smooth measurement errors, are introduced in Section 2; the mixed kernel density estimators of the joint density function, as well their derivatives, are presented in Section 3. The mean squared error (MSE) of the proposed mixed kernel estimator, the optimal convergence rates of the nonparametric estimators for the density functions and their derivatives, as well as its asymptotic normality will be discussed in Section 4. A simulation study is conducted in Section 5 to evaluate the finite sample performance of the proposed estimators. All the technical proofs are deferred to the appendix.

2.2 Notations and Conditions

Suppose that a random triple $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})$ with $\boldsymbol{Y} \in \mathbb{R}^d$, $\boldsymbol{X} \in \mathbb{R}^p$, $\boldsymbol{Z} \in \mathbb{R}^q$, respectively, $d \geq 1$, $p \geq 1$ and $q \geq 1$ is of interest in a statistical model. Observations can be directly made on $\boldsymbol{Y}$. However, due to the imperfection on measurement instruments or human errors, we cannot observe $\boldsymbol{X}$ and $\boldsymbol{Z}$ directly. Instead, we can observe $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$, $\boldsymbol{S} = \boldsymbol{Z} + \boldsymbol{V}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the measurement errors. Throughout the paper, we shall assume that $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})$ are independent of $(\boldsymbol{U}, \boldsymbol{V})$, $\boldsymbol{U}$ is independent of $\boldsymbol{V}$. Furthermore, for the sake of identifiability, the density functions of $\boldsymbol{U}$ and $\boldsymbol{V}$ are assumed to be known. Although replicated observations or other assumptions on $\boldsymbol{X}$ and $\boldsymbol{Z}$ can help us relax this rigid requirement, we will not pursue this direction in the current paper.

To facilitate the presentation, we begin with a relatively complete list of notations, including the multiple index notations particularly designed for dealing with the multivariate functions.

- Regular capital letters, for example, X, Y, denote random variables, and bold capital letters, for example, $\boldsymbol{X}, \boldsymbol{Y}$, denote random vectors;

- Regular lower case letters, for example, x, y, denote non-random variables, and bold lower case letters, for example, $\boldsymbol{x}, \boldsymbol{y}$, denote non-random vectors;

- The density function of $\boldsymbol{X}$ or X is denoted by $f_{\boldsymbol{X}}$, f_X, respectively;

- For any real vectors $\boldsymbol{a} = (a_1, \ldots, a_k)^T$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^T$, denote $[\boldsymbol{a}] = a_1 + a_2 + \cdots + a_k$,

$$\boldsymbol{a}_{\times}^{\boldsymbol{\alpha}} = a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_k^{\alpha_k}, \quad |\boldsymbol{a}|_{\times}^{\boldsymbol{\alpha}} = |a_1|^{\alpha_1} |a_2|^{\alpha_2} \cdots |a_k|^{\alpha_k}, \quad \boldsymbol{a}! = a_1! a_2! \cdots a_k!,$$
$$\boldsymbol{a}_{+}^{\boldsymbol{\alpha}} = a_1^{\alpha_1} + a_2^{\alpha_2} + \cdots + a_k^{\alpha_k}, \quad |\boldsymbol{a}|_{+}^{\boldsymbol{\alpha}} = |a_1|^{\alpha_1} + |a_2|^{\alpha_2} + \cdots + |a_k|^{\alpha_k};$$

- For vectors $\boldsymbol{a}_1 = (a_{11}, \ldots, a_{1d})^T$, $\boldsymbol{a}_2 = (a_{21}, \ldots, a_{2d})^T$, and a real number c,

$$\boldsymbol{a}_1 + \boldsymbol{a}_2 = (a_{11} + a_{21}, \ldots, a_{1d} + a_{2d})^T, \quad \boldsymbol{a}_1 + c = (a_{11} + c, \ldots, a_{1d} + c)^T,$$
$$\boldsymbol{a}_1 \boldsymbol{a}_2 = (a_{11} a_{21}, a_{12} a_{22}, \ldots, a_{1d} a_{2d})^T;$$

- For any multivariate function $f(\boldsymbol{y}, \boldsymbol{x})$, $\boldsymbol{y} = (y_1, \ldots, y_d)^T$, $\boldsymbol{x} = (x_1, \ldots, x_p)^T$, and $\boldsymbol{a} = (a_1, \ldots, a_d)^T$, $\boldsymbol{b} = (b_1, \ldots, b_p)^T$, $\boldsymbol{\alpha} = (\boldsymbol{a}^T, \boldsymbol{b}^T)^T$, $a_i, b_j \in \{0, 1, \ldots\}$ for $i = 1, \ldots, d$ and $j = 1, \ldots, p$,

$$f^{(\boldsymbol{\alpha})}(\boldsymbol{y}, \boldsymbol{x}) = \frac{\partial^{[\boldsymbol{\alpha}]} f(\boldsymbol{y}, \boldsymbol{x})}{\partial \boldsymbol{y}^{\boldsymbol{a}} \partial \boldsymbol{x}^{\boldsymbol{b}}} = \frac{\partial^{a_1 + \cdots + a_d + b_1 + \cdots + b_p} f(\boldsymbol{y}, \boldsymbol{x})}{\partial y_1^{a_1} \cdots \partial y_d^{a_d} \partial x_1^{b_1} \cdots \partial x_p^{b_p}};$$

- For a random variable X or vector $\boldsymbol{X}$, its characteristic function is denoted by $\phi_X(t)$ or $\phi_{\boldsymbol{X}}(\boldsymbol{t})$;

- A d-dimensional vector of all 1's is denoted by $\boldsymbol{J}_d$, and $\mathbf{1}_{dj}$ denotes a d-dimensional vector of 0's except for the j-th element being 1;

– Denote $\boldsymbol{D} = (\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})$, $(\boldsymbol{Y}, \boldsymbol{X})$ or $(\boldsymbol{Y}, \boldsymbol{Z})$ depending on whether $\boldsymbol{Z}$, $\boldsymbol{X}$ present or not.

Based on how the tails of the characteristic functions decay to 0, two types of measurement errors are defined in Fan (1991a). The ordinary smooth measurement errors have characteristic functions decaying to 0 at an algebraic rate, and the super-smooth measurement errors have characteristic functions vanishing at an exponential rate. To be specific, in our current setup, we shall assume that the components of the measurement errors of the same type are independent, and

– $\boldsymbol{U} = (U_1, \ldots, U_p)$ is ordinary smooth. That is, $U_1, \ldots, U_p$ are independent, and for $j = 1, \ldots, p$, and $|t|$ large enough, $\tau_{j0}|t|^{-\beta_j} \le |\phi_{U_j}(t)| \le \tau_{j1}|t|^{-\beta_j}$, where τ_{j0}, τ_{j1} and β_j are positive numbers. A typical example of the ordinary smooth measurement error is the double exponential random variable.

– $\boldsymbol{V} = (V_1, \ldots, V_q)$ is super smooth. That is, $V_1, \ldots, V_q$ are independent, and for $j = 1, \ldots, q$, and $|t|$ large enough, $\zeta_{j0}|t|^{\mu_{j0}} \exp(-|t|^{\gamma_j}/\xi_j) \le |\phi_{V_j}(t)| \le \zeta_{j1}|t|^{\mu_{j1}} \exp(-|t|^{\gamma_j}/\xi_j)$, where ζ_{j0}, ζ_{j1}, γ_j, ξ_j are positive numbers and μ_{j0}, μ_{j1} are real numbers. A typical example of the super-smooth measurement error is the normal random variable.

Here we assume that the components of the measurement errors are independent. It might be more general to investigate the measurement errors with some dependent structures. Indeed, although the proposed estimation procedure is designed for cases with independent measurement error, it can well accommodate some dependent scenarios. For example, if $\boldsymbol{V} \sim N_q(\boldsymbol{0}, \Sigma)$, then by transformation $\Sigma^{-1/2}\boldsymbol{S} = \Sigma^{-1/2}\boldsymbol{Z} + \Sigma^{-1/2}\boldsymbol{V}$, we obtain a multivariate standard normal measurement error $\Sigma^{-1/2}\boldsymbol{V}$ whose components are independent. Thus a density estimator for $\boldsymbol{Z}$ can be constructed from any estimator for the density function of $\Sigma^{-1/2}\boldsymbol{Z}$. However, for other measurement errors such as the multivariate Laplace distribution, variable transforming may not change a dependent structure into an independent structure. In particular, there are at least three different definitions for the multivariate Laplace distribution, see McGraw and Wagner (1968), Johnson and Kotz (1972), Johnson

(1987), Anderson (1992), Osiewalski and Steel (1993), Marshall and Olkin (1993) and Kotz et al. (2001) for more details. According to Anderson (1992), a random vector, say $\boldsymbol{U}$, is said to have a multivariate Laplace distribution if its characteristic function is given be $(1+\boldsymbol{t}'\Sigma\boldsymbol{t}/2)^{-1}$. Clearly, the transformation $\Sigma^{-1/2}\boldsymbol{U}$ cannot generate a random vector with independent components. So, for the sake of simplicity and not losing our focus on the main research topic, throughout this paper, we shall assume that all the components in the measurement errors are independent, but may have different smoothness.

It is well known that the smoothness of the measurement error has a significant effects on the convergence rate of the nonparametric estimate of the density estimates and regression function estimates. Literature shows that estimating the density function of a variable measured with ordinary smooth measurement errors amounts to estimating the derivative of the density functions of a variable free of measurement error, while it is more difficult to estimate such quantities when the measurement errors are super-smooth. See Fan (1991a), Fan (1991b), Fan and Truong (1993), Shi et al. (2020) and the references therein for more details.

2.3 Mixed Kernel Density Estimators

The idea of constructing the mixed kernel estimator in Shi et al. (2020) can be readily extended to construct the estimator for the density function $f_{\boldsymbol{D}}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$, as well as its derivatives. Denoted by $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$, the mixed kernel density estimator for the triple $\boldsymbol{D}$ is defined as

$$\hat{f}_{\boldsymbol{D}}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) = \frac{1}{n\boldsymbol{h}_{\times}^{\boldsymbol{J}_d}\boldsymbol{\eta}_{\times}^{\boldsymbol{J}_p}\boldsymbol{\rho}_{\times}^{\boldsymbol{J}_q}}\cdot \sum_{j=1}^{n}\left[\prod_{l_1=1}^{d} L\left(\frac{y_{l_1}-Y_{jl_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}\left(\frac{x_{l_2}-W_{jl_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K_{nl_3}\left(\frac{z_{l_3}-S_{jl_3}}{\rho_{l_3}}\right)\right],$$

where the elements in vectors $\boldsymbol{h}=(h_1,\ldots,h_d)^T$, $\boldsymbol{\eta}=(\eta_1,\ldots,\eta_p)^T$, $\boldsymbol{\rho}=(\rho_1,\ldots,\rho_q)^T$ are sequence of positive numbers tending to 0 as the sample size $n\to\infty$. $\boldsymbol{Y}_j=(Y_{j1},\ldots,Y_{jd})^T$,

$\boldsymbol{W}_j = (W_{j1}, \ldots, W_{jp})^T$, $\boldsymbol{S}_j = (S_{j1}, \ldots, S_{jq})^T$, $j = 1, 2, \ldots, n$, are samples from $\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{S}$, respectively. L is a classical univariate kernel density function. For $l_2 = 1, 2, \ldots, p$, $l_3 = 1, 2, \ldots, q$, H_{nl_2} and K_{nl_3} are univariate deconvolution kernel functions defined by

$$H_{nl_2}(x) = \frac{1}{2\pi} \int e^{-\mathbf{i}tx} \frac{\phi_H(t)}{\phi_{U_{l_2}}(t/\eta)} dt, \quad K_{nl_3}(z) = \frac{1}{2\pi} \int e^{-\mathbf{i}tz} \frac{\phi_K(t)}{\phi_{V_{l_3}}(t/\rho)} dt$$

with H and K being two univariate classical kernel density functions. It is well known that the kernel has little effect on the performance of the estimator, so for the sake of simplicity, we choose the kernel functions to be the same for each type of measurement error.

Let $\boldsymbol{k} = (\boldsymbol{a}^T, \boldsymbol{b}^T, \boldsymbol{c}^T)$, where the entries in $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are pre-specified nonnegative integers, and define

$$H_{nl_2}^{(b_{l_2})}(x) = \frac{1}{2\pi} \int (-\mathbf{i}t)^{b_{l_2}} e^{-\mathbf{i}tx} \frac{\phi_H(t)}{\phi_{U_{l_2}}(t/\eta)} dt, \quad K_{nl_3}^{(c_{l_3})}(z) = \frac{1}{2\pi} \int (-\mathbf{i}t)^{c_{l_3}} e^{-\mathbf{i}tz} \frac{\phi_K(t)}{\phi_{V_{l_3}}(t/\rho)} dt,$$

where $l_2 = 1, 2, \ldots, p$ and $l_3 = 1, 2, \ldots, q$. Then we define a mixed kernel estimator of the $\boldsymbol{k}$-th derivative of $f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$, $\partial^{[\boldsymbol{k}]} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) / \partial \boldsymbol{y}^{\boldsymbol{a}} \partial \boldsymbol{x}^{\boldsymbol{b}} \partial \boldsymbol{z}^{\boldsymbol{c}}$, as

$$\begin{aligned} \hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) = & \frac{1}{n \boldsymbol{h}_{\times}^{\boldsymbol{a}+1} \boldsymbol{\eta}_{\times}^{\boldsymbol{b}+1} \boldsymbol{\rho}_{\times}^{\boldsymbol{c}+1}} \cdot \\ & \sum_{j=1}^{n} \left[\prod_{l_1=1}^{d} L^{(a_{l_1})} \left(\frac{y_{l_1} - Y_{jl_1}}{h_{l_1}} \right) \prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})} \left(\frac{x_{l_2} - W_{jl_2}}{\eta_{l_2}} \right) \prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})} \left(\frac{z_{l_3} - S_{jl_3}}{\rho_{l_3}} \right) \right]. \end{aligned}$$

One can certainly consider the possibility of using other methods to construct the estimators for the density functions and their derivatives. For example, the truncated deconvolution kernel proposed in Liu and Taylor (1989) and the truncated orthogonal series estimator introduced in Efromovich (1997) etc. could be the candidates for dealing with the measurement error components in the mixed types of data. However, in this paper we only focus on the estimators constructed using the classical deconvolution kernel.

2.4 Main Results

To find the optimal convergence rate among all nonparametric estimators for the density function $f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ and its derivatives, the density function to be estimated is assumed to be a member from a distribution family. Let α be a positive real number in $(0, 1)$, m be a positive integer, and B be a positive real number. We assume that the density function $f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ belongs to the following distribution family

$$\begin{aligned}\mathscr{C}_{\alpha,m,B} = \Bigg\{ f_{\boldsymbol{D}} : &\left| \frac{\partial^m f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}}} \right|_{\substack{\boldsymbol{y}=\boldsymbol{y}^* \\ \boldsymbol{x}=\boldsymbol{x}^* \\ \boldsymbol{z}=\boldsymbol{z}^*}} - \frac{\partial^m f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}}} \Bigg| \\ &\leq B(|\boldsymbol{y}^* - \boldsymbol{y}|_+^\alpha + |\boldsymbol{x}^* - \boldsymbol{x}|_+^\alpha + |\boldsymbol{z}^* - \boldsymbol{z}|_+^\alpha), \\ &\text{for all positive integers } \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}} \text{ s.t. } [\tilde{\boldsymbol{a}}] + [\tilde{\boldsymbol{b}}] + [\tilde{\boldsymbol{c}}] = m \Bigg\}. \end{aligned} \tag{2.4.1}$$

That is, all the m-th order derivatives of $f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ are α-order Hölder continuous. However, to show the asymptotic normality of the proposed mixed kernel estimators, the constraint in the family $\mathscr{C}_{\alpha,m,B}$ appears too strong. In fact, the conditions (O1) and (S1) below are already sufficient for this purpose. The following list contains the technical conditions needed to present and prove the theoretical results.

(C1). For all $\tilde{\boldsymbol{a}} \leq \boldsymbol{a}$, $\tilde{\boldsymbol{b}} \leq \boldsymbol{b}$, $\tilde{\boldsymbol{c}} \leq \boldsymbol{c}$, $\tilde{\boldsymbol{k}} = (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}})$, and all $l_1 \leq \max\{\tilde{a}_j : j = 1, 2, \ldots, d\}$, $l_2 \leq \max\{\tilde{b}_j : j = 1, 2, \ldots, p\}$, $l_3 \leq \max\{\tilde{c}_j : j = 1, 2, \ldots, q\}$,

$$\frac{\partial^{[\tilde{\boldsymbol{k}}]} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}}} \to 0, \quad L^{(l_1)}(y) \to 0, \quad H^{(l_2)}(x) \to 0, \quad K^{(l_3)}(z) \to 0,$$

as $|\boldsymbol{y}| + |\boldsymbol{x}| + |\boldsymbol{z}| \to \infty$, $|y|, |x|, |z| \to \infty$, respectively.

(C2). We assume that L, H, K are $(m - [\boldsymbol{k}] + 1)$-th order kernel function. That is, for

$\tilde{\boldsymbol{a}}$, $\tilde{\boldsymbol{b}}$ and $\tilde{\boldsymbol{c}}$ such that $[\tilde{\boldsymbol{a}}]+[\tilde{\boldsymbol{b}}]+[\tilde{\boldsymbol{c}}] \leq m-[\boldsymbol{k}]+1$, $m \geq [\boldsymbol{k}]$,

$$\iiint \boldsymbol{y}^{\tilde{\boldsymbol{a}}} \boldsymbol{x}^{\tilde{\boldsymbol{b}}} \boldsymbol{z}^{\tilde{\boldsymbol{c}}} \prod_{l_1=1}^{d} L(y_{l_1}) \prod_{l_2=1}^{p} H(x_{l_2}) \prod_{l_3=1}^{q} K(z_{l_3}) d\boldsymbol{y} d\boldsymbol{x} d\boldsymbol{z} = 0.$$

(C3). The marginal density function of $\boldsymbol{Y}, \boldsymbol{X}$ is bounded.

(C4). For the ordinary-smooth error $\boldsymbol{V}$, we assume that for $j = 1, 2, \ldots, p$, $|t^{-\beta_j - l} \phi_{V_j}^{(l)}(t)| \leq d_l$ as $|t| \to \infty$, where $d_l, l = 0, 1, 2$, are positive constants.

(C5). For the super-smooth error $\boldsymbol{U}$, we assume that for $j = 1, 2, \ldots, q$, $P(|U_j - u| \leq |u|^{\mu_0}) = O(|u|^{-(\mu-\mu_0)})$ as $|u| \to \infty$ for some $\mu_0 \in (0, 1)$ and $\mu > 1 + \mu_0$.

(L1). The kernel function L has a characteristic function ϕ_L such that for $l_1 = 1, 2, \ldots, d$,

$$\int |s_{l_1}|^{2a_{l_1}} |\phi_L(s_{l_1})| ds_{l_1} < \infty.$$

(O1). $f_{\boldsymbol{D}}^{(m)}(\boldsymbol{y}, \boldsymbol{x})$ exists and is continuous.

(O2). For each $l_2 = 1, 2, \ldots, p$, the characteristic function $\phi_{U_{l_2}}$ satisfies $\phi_{U_{l_2}}(t) t^{\beta_{l_2}} \to c_{l_2}$ and $\phi'_{U_{l_2}}(t) t^{\beta_{l_2}+1} \to -\beta_{l_2} c_{l_2}$ as $t \to \infty$, where $c_{l_2} \neq 0$, $\beta_{l_2} \geq 0$. Also, $\phi_{U_{l_2}}(t) \neq 0$ for all t.

(O3). The kernel function H has a symmetric characteristic function, having $m+2$ bounded integrable derivatives, and $\phi_H(t) = 1 + O(|t|^m)$ as $t \to 0$. Also,

$$\int [|\phi_H(t)| + |\phi'_H(t)|] |t|^{\beta_{l_2} + b_{l_2}} dt < \infty, \quad \int |\phi_H(t)|^2 |t|^{2\beta_{l_2} + 2b_{l_2}} dt < \infty.$$

(S1). $f_{\boldsymbol{D}}^{(m)}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ or $f_{\boldsymbol{D}}^{(m)}(\boldsymbol{y}, \boldsymbol{z})$ exists and is continuous.

(S2). The characteristic function $\phi_K(t)$ is symmetric and supported on $[-1, 1]$ and having the first $m+2$ continuous derivatives, and also $\phi_K(t) > c_3(1-t)^{m+3}$ for $t \in [1-\delta, 1]$, where δ is a sufficiently small number and c_3 is a positive number.

(S3). For the bandwidth $\boldsymbol{h}, \boldsymbol{\eta}, \boldsymbol{\rho}$, $n\boldsymbol{h}_\times \boldsymbol{\eta}_\times \to \infty$, and $\rho_{l_3} = O((\log(n\boldsymbol{\eta}_\times \boldsymbol{h}_\times))^{-1/\beta_{l_3}})$ for $l_3 = 1, 2, \ldots, q$, as $n \to \infty$.

Conditions (C1)-(C3) are needed to find the MSE of the proposed estimator. Note that $C1$ is a much stronger condition than we actually need, but we adopt (C1) here simply because a less strict condition appears more complicated in notations. (C3) is used for deriving the order of the variance of the proposed estimator. One can further assume that $f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ is bounded to achieve a higher order, but the resulting order is not high enough to change the overall convergence rate. (C4) and (C5) are needed for deriving the optimal convergence rate of all nonparametric estimators. Conditions (L1), (O1)-(O3) imposed on the ordinary smooth components and conditions (S1)-(S3) imposed on the super-smooth components are required to show the asymptotic normality.

The following theorem presents the MSE of the proposed estimator.

Theorem 2.4.1. *Suppose (C1), (C2) and (C3) hold. Then*

$$\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}^{(\boldsymbol{k})}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) - f^{(\boldsymbol{k})}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})\right]^2 \tag{2.4.2}$$

$$= O\Bigg(\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_1=1}^{d} B_{j_1,\tilde{\boldsymbol{k}},1} |\boldsymbol{h}|_\times^{\tilde{\boldsymbol{a}}+\alpha \mathbf{1}_{dj_1}} |\boldsymbol{\eta}|_\times^{\tilde{\boldsymbol{b}}} |\boldsymbol{\rho}|_\times^{\tilde{\boldsymbol{c}}} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_2=1}^{p} B_{j_2,\tilde{\boldsymbol{k}},2} |\boldsymbol{h}|_\times^{\tilde{\boldsymbol{a}}} |\boldsymbol{\eta}|_\times^{\tilde{\boldsymbol{b}}+\alpha \mathbf{1}_{pj_2}} |\boldsymbol{\rho}|_\times^{\tilde{\boldsymbol{c}}} +$$

$$\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_3=1}^{q} B_{j_3,\tilde{\boldsymbol{k}},3} |\boldsymbol{h}|_\times^{\tilde{\boldsymbol{a}}} |\boldsymbol{\eta}|_\times^{\tilde{\boldsymbol{b}}} |\boldsymbol{\rho}|_\times^{\tilde{\boldsymbol{c}}+\alpha \mathbf{1}_{qj_3}} \Bigg)^2 + O\left(\frac{1}{n \boldsymbol{h}_\times^{2\boldsymbol{a}+1} \boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1} \boldsymbol{\rho}_\times^{2\boldsymbol{c}-2\tilde{\boldsymbol{b}}+2}} \exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}} \boldsymbol{\xi}^{-1})_+) \right).$$

where

$$B_{j_1,\tilde{\boldsymbol{k}},1} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{b}}!\tilde{\boldsymbol{c}}!} \iiint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| |\tilde{\boldsymbol{y}}|_\times^{\tilde{\boldsymbol{a}}+\alpha \mathbf{1}_{dj_1}} |\tilde{\boldsymbol{x}}|_\times^{\tilde{\boldsymbol{b}}} |\tilde{\boldsymbol{z}}|_\times^{\tilde{\boldsymbol{c}}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}},$$

$$B_{j_2,\tilde{\boldsymbol{k}},2} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{b}}!\tilde{\boldsymbol{c}}!} \iiint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| |\tilde{\boldsymbol{y}}|_\times^{\tilde{\boldsymbol{a}}} |\tilde{\boldsymbol{x}}|_\times^{\tilde{\boldsymbol{b}}+\alpha \mathbf{1}_{pj_2}} |\tilde{\boldsymbol{z}}|_\times^{\tilde{\boldsymbol{c}}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}},$$

$$B_{j_3,\tilde{\boldsymbol{k}},3} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{b}}!\tilde{\boldsymbol{c}}!} \iiint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| |\tilde{\boldsymbol{y}}|_\times^{\tilde{\boldsymbol{a}}} |\tilde{\boldsymbol{x}}|_\times^{\tilde{\boldsymbol{b}}} |\tilde{\boldsymbol{z}}|_\times^{\tilde{\boldsymbol{c}}+\alpha \mathbf{1}_{qj_3}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}}$$

for $j_1 = 1, 2, \ldots, d$, $j_2 = 1, 2, \ldots, p$, $j_3 = 1, 2, \ldots, q$ *and* $\bar{\boldsymbol{b}} = (\bar{b}_1, \ldots, \bar{b}_q)$ *and* $\bar{b}_j = 0$ *if* $\mu_{j0} \geq 0$ *and* μ_{j0} *if* $\mu_{j0} < 0$, *for* $j = 1, 2, \ldots, q$.

The MSEs in some special cases can be easily derived from Theorem 2.4.1. If equal bandwidths are chosen for each type of random variables. That is, assuming that $h_1 = \cdots = h_d = h$, $\eta_1 = \cdots = \eta_p = \eta$, $\rho_1 = \cdots = \rho_q = \rho$, (2.4.2) becomes

$$\begin{aligned}
&\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})\right]^2 \qquad (2.4.3)\\
&= O\Bigg(\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} B_{\tilde{\boldsymbol{k}},1} h^{[\tilde{\boldsymbol{a}}]+\alpha} \eta^{[\tilde{\boldsymbol{b}}]} \rho^{[\tilde{\boldsymbol{c}}]} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} B_{\tilde{\boldsymbol{k}},2} h^{[\tilde{\boldsymbol{a}}]} \eta^{[\tilde{\boldsymbol{b}}]+\alpha} \rho^{[\tilde{\boldsymbol{c}}]} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} B_{\tilde{\boldsymbol{k}},3} h^{[\tilde{\boldsymbol{a}}]} \eta^{[\tilde{\boldsymbol{b}}]} \rho^{[\tilde{\boldsymbol{c}}]+\alpha} \Bigg)^2 \\
&\quad + O\left(\frac{1}{n h^{2[\boldsymbol{a}]+d} \eta^{2[\boldsymbol{b}]+2[\boldsymbol{\beta}]+p} \rho^{2[\boldsymbol{c}]-2[\bar{\boldsymbol{b}}]+2q}} \exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}} \boldsymbol{\xi}^{-1})_+) \right),
\end{aligned}$$

where $\tilde{\boldsymbol{k}} = (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}})$, and $B_{\tilde{\boldsymbol{k}},1} = \sum_{j_1=1}^{d} B_{j_1,\tilde{\boldsymbol{k}},1}$, $B_{\tilde{\boldsymbol{k}},2} = \sum_{j_2=1}^{p} B_{j_2,\tilde{\boldsymbol{k}},2}$, $B_{\tilde{\boldsymbol{k}},3} = \sum_{j_3=1}^{q} B_{j_3,\tilde{\boldsymbol{k}},3}$. If we further assume that all the elements in $\boldsymbol{h}, \boldsymbol{\eta}, \boldsymbol{\rho}$ are the same, say all equal to h, then we have

$$\begin{aligned}
&\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})\right]^2 \qquad (2.4.4)\\
&= O\left(h^{2m-2[\boldsymbol{k}]+2\alpha}\right) + O\left(\frac{1}{n h^{2[\boldsymbol{k}]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q}} \exp((2\boldsymbol{h}^{-\boldsymbol{\gamma}} \boldsymbol{\xi}^{-1})_+) \right).
\end{aligned}$$

On the other hand, if not all types of measurement-error contaminated variables present in the data, for example, if there is no super-smooth component, then (2.4.2) becomes

$$\begin{aligned}
&\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x})\right]^2 \qquad (2.4.5)\\
&= O\Bigg(\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_1=1}^{d} B_{j_1,\tilde{\boldsymbol{k}},1} |\boldsymbol{h}|_{\times}^{\tilde{\boldsymbol{a}}+\alpha \boldsymbol{1}_{dj_1}} |\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_2=1}^{p} B_{j_2,\tilde{\boldsymbol{k}},2} |\boldsymbol{h}|_{\times}^{\tilde{\boldsymbol{a}}} |\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}+\alpha \boldsymbol{1}_{pj_2}} \Bigg)^2 \\
&\quad + O\left(\frac{1}{n \boldsymbol{h}_{\times}^{2\boldsymbol{a}+1} \boldsymbol{\eta}_{\times}^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}} \right).
\end{aligned}$$

where

$$B_{j_1,\tilde{\boldsymbol{k}},1} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{b}}!} \iint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| |\tilde{\boldsymbol{y}}|_{\times}^{\tilde{\boldsymbol{a}}+\alpha \mathbf{1}_{dj_1}} |\tilde{\boldsymbol{x}}|_{\times}^{\tilde{\boldsymbol{b}}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}},$$

$$B_{j_2,\tilde{\boldsymbol{k}},2} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{b}}!} \iint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| |\tilde{\boldsymbol{y}}|_{\times}^{\tilde{\boldsymbol{a}}} |\tilde{\boldsymbol{x}}|_{\times}^{\tilde{\boldsymbol{b}}+\alpha \mathbf{1}_{pj_2}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}}.$$

If there is no ordinary smooth component, then (2.4.2) becomes

$$\begin{aligned} &\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{z})\right]^2 \\ &= O\Bigg(\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_1=1}^{d} B_{j_1,\tilde{\boldsymbol{k}},1} |\boldsymbol{h}|_{\times}^{\tilde{\boldsymbol{a}}+\alpha \mathbf{1}_{dj_1}} |\boldsymbol{\rho}|_{\times}^{\tilde{\mathbf{c}}} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_3=1}^{q} B_{j_3,\tilde{\boldsymbol{k}},3} |\boldsymbol{h}|_{\times}^{\tilde{\boldsymbol{a}}} |\boldsymbol{\rho}|_{\times}^{\tilde{\mathbf{c}}+\alpha \mathbf{1}_{qj_3}} \Bigg)^2 \\ &\quad + O\left(\frac{1}{n \boldsymbol{h}_{\times}^{2\boldsymbol{a}+1} \boldsymbol{\rho}_{\times}^{2\boldsymbol{c}-2\tilde{\boldsymbol{b}}+2}} \exp((2\boldsymbol{\rho}^{-\gamma}\boldsymbol{\xi}^{-1})_{+}) \right), \end{aligned} \tag{2.4.6}$$

where

$$B_{j_1,\tilde{\boldsymbol{k}},1} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{c}}!} \iint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| |\tilde{\boldsymbol{y}}|_{\times}^{\tilde{\boldsymbol{a}}+\alpha \mathbf{1}_{dj_1}} |\tilde{\boldsymbol{z}}|_{\times}^{\tilde{\mathbf{c}}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{z}},$$

$$B_{j_3,\tilde{\boldsymbol{k}},3} = \frac{B}{\tilde{\boldsymbol{a}}!\tilde{\boldsymbol{c}}!} \iint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| |\tilde{\boldsymbol{y}}|_{\times}^{\tilde{\boldsymbol{a}}} |\tilde{\boldsymbol{z}}|_{\times}^{\tilde{\mathbf{c}}+\alpha \mathbf{1}_{qj_3}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{z}}.$$

Given the complexity of (2.4.2), it is rather difficult to consider the optimal convergence rate of the estimator $\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$, even for the special case (2.4.3). In the following, we shall focus on (2.4.4), that is, a single bandwidth, say h, is applied for all components in the model. In this case, by (2.4.4), the MSE has the form of

$$\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\right]^2 \tag{2.4.7}$$
$$= O\left(h^{2m-2[\boldsymbol{k}]+2\alpha}\right) + O\left(\frac{1}{nh^{2[\boldsymbol{k}]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q}} \exp\left(2\sum_{j=1}^{q}\frac{1}{\xi_j h^{\gamma_j}}\right)\right).$$

Without loss of generality, let's assume that $\gamma_1 = \max\{\gamma_j : j = 1, 2, \ldots, q\}$. To guarantee the variance part goes to 0, we have to make sure that h cannot go to 0 faster than $(\log n)^{-1/\gamma_1}$. To show this, let $h = a_n(\log n)^{-1/\gamma_1}$. If $a_n \to 0$ as $n \to \infty$, then

$$\frac{1}{nh^{2[k]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q}} \exp\left(\sum_{j=1}^{q}\frac{2}{\xi_j h^{\gamma_j}}\right)$$
$$= \frac{(\log n)^{\{2[k]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q\}/\gamma_1}}{n^{1-2/(\xi_1 a_n^{\gamma_1})} a_n^{2[k]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q}} \exp\left[2\sum_{j=2}^{q}\frac{(\log n)^{\gamma_j/\gamma_1}}{\xi_j a_n^{\gamma_j}}\right]$$

goes to ∞. On the other hand, a_n cannot go to ∞ faster than $(\log n)^{1/\gamma_1}$, otherwise, the bias part cannot be controlled. Choosing a_n be any constant a such that $1-2(\sum_{j=1}^{q}\xi_j a^{\gamma_j})^{-1} > 0$, and noting that the right-hand side of the above equality is bounded above by

$$\frac{c(\log n)^{\{2[k]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q\}/\gamma_1}}{n^{1-2\sum_{j=1}^{q}(\xi_j a^{\gamma_j})^{-1}} a^{2[k]+2[\boldsymbol{\beta}]-2[\bar{\boldsymbol{b}}]+d+p+2q}}$$

for some positive constant c. It is easy to see that this bound converges to 0 as $n \to \infty$. Therefore, we have

Theorem 2.4.2. *Under the conditions (C1-C3), if we choose $h = O\left((\log n)^{-1/\gamma_1}\right)$, then*

$$\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\right]^2 = O\left((\log n)^{-(2m-2[\boldsymbol{k}]+2\alpha)/\gamma_1}\right).$$

If there is no super-smooth part, then from (2.4.5), we have

$$\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x})\right]^2 = O(h^{2m-2[\boldsymbol{k}]+2\alpha}) + O\left(\frac{1}{nh^{2[\boldsymbol{k}]+2[\boldsymbol{\beta}]+d+p}}\right),$$

where $[\boldsymbol{k}] = [\boldsymbol{a}] + [\boldsymbol{b}]$. In this case, an optimal bandwidth h, after neglecting some constants not depending on the sample size, can be obtained by minimizing the sum $h^{2m-2[\boldsymbol{k}]+2\alpha} + n^{-1}h^{-2[\boldsymbol{k}]-2[\boldsymbol{\beta}]-d-p}$. Therefore, we have

Theorem 2.4.3. *Suppose (C1)-(C3) holds for $f_{\boldsymbol{D}}(\boldsymbol{y},\boldsymbol{x})$ and the kernel functions L, H. If we choose $h = O(n^{-1/(2m+2\alpha+2[\boldsymbol{\beta}]+d+p)})$, then*

$$\sup_{f_{\boldsymbol{D}} \in \mathscr{C}_{m,\alpha,B}} E\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x})\right]^2 = O\left(n^{-(2m-2[\boldsymbol{k}]+2\alpha)/(2m+2[\boldsymbol{\beta}]+2\alpha+d+p)}\right).$$

Theorems 2.4.2 shows that when the super-smooth measurement errors present in the data, then the convergence rate of the proposed mixed kernel density estimator of $f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$ is rather slow, and it is the smoothest component, with the largest γ-values in the super-smooth measurement errors, that determines the convergence rate. If there is no super-smooth measurement error presents in the data, from Theorem 2.4.3, the convergence rate is significantly increase to the scenarios as if we are estimating the derivatives of a density function when all the data can be observed directly. In both cases, the variables of no measurement errors cannot control the convergence rate.

In the univariate case, Fan (1991a) established the optimal convergence rates among the class of nonparametric estimators of the derivatives of the density functions for both the ordinary and super-smooth variables, and also proved that the deconvolution kernel estimators achieve the optimal convergence rate. In the bivariate cases where one variable is free of measurement error and another variable is contaminated with either one of the ordinary or super-smooth measurement error, Shi et al. (2020) also derived the optimal convergence rates for nonparametric estimators of the derivatives of the joint density functions, and the mixed kernel estimators indeed attain the optimal convergence rate.

The question of whether the proposed mixed kernel estimator achieving the optimal convergence rate is answered by the following theorem.

Theorem 2.4.4. *Suppose all the conditions (C1)-(C5) hold. Then the proposed mixed kernel estimators, with equal bandwidth $h = O((\log n)^{-1/\gamma_1})$ when the super-smooth measurement error presents, and $h = O(n^{-1/(2m+2\alpha+2[\boldsymbol{\beta}]+d+p)})$ when the super-smooth measurement error does not present, attain the optimal convergence rates.*

Based on Theorem 2.4.4, we have the following interesting discoveries: (1). Choosing different bandwidths for different components cannot improve the convergence rate; (2). When the super-smooth measurement error appears in the data, the optimal convergence rate depends on the dimensions of $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}$ only through $[\boldsymbol{k}]$, the order of derivatives, however, when only the ordinary-smooth measurement error and error-free variables appear in the data, then the optimal convergence rate not only depends on $[\boldsymbol{k}]$, also depends on the dimensions d and p explicitly. To see the significance of the latter, suppose we are interested in the density estimator itself only, that is $[\boldsymbol{k}] = 0$, then we see that, when super-smooth measurement error appears in the data, the optimal convergence rate is free of dimensions, while this is not the case when it does not show in the data.

From Theorem 2.4.4, we see that the convergence rate are very different when the super-smooth component shows or does not show up in the data. The following theorem presents the asymptotic result for the mixed kernel estimator when the super-smooth is absent from the data.

Theorem 2.4.5. *Under conditions (L1), (O1), (O2) and (O3), we have*

$$\frac{\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}) - E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x})}{\sqrt{Var(\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}))}} \Longrightarrow N(0, 1).$$

Moreover, if we further have

$$\sqrt{n\boldsymbol{h}_{\times}^{2\boldsymbol{a}+1}\boldsymbol{\eta}_{\times}^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}}\left(\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]}|\boldsymbol{h}|_{\times}^{\tilde{\boldsymbol{a}}}|\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}}\right)[\boldsymbol{h}_{+}^{\alpha\boldsymbol{1}_d} + \boldsymbol{\eta}_{+}^{\alpha\boldsymbol{1}_p}] \to 0,$$

then

$$\sqrt{n\boldsymbol{h}_{\times}^{2\boldsymbol{a}+1}\boldsymbol{\eta}_{\times}^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}V^{-1}(\boldsymbol{y},\boldsymbol{x})}\left[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x})-f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x})\right]\Longrightarrow N(0,1),$$

where

$$V(\boldsymbol{y},\boldsymbol{x})=\frac{f_{\boldsymbol{Y},\boldsymbol{W}}(\boldsymbol{y},\boldsymbol{x})}{(2\pi)^{d+p}C^2}\iint\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}|s_{l_1}|^{2a_{l_1}}|t_{l_2}|^{2b_{l_2}+2\beta_{l_2}}|\phi_L(s_{l_1})|^2|\phi_H(t_{l_2})|^2d\boldsymbol{s}d\boldsymbol{t}. \tag{2.4.8}$$

Note that $\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x})$ can be written as the average of ξ_{nj}, $j=1,2,\ldots,n$, where

$$\xi_{nj}=\frac{1}{\boldsymbol{h}_{\times}^{\boldsymbol{a}+1}\boldsymbol{\eta}_{\times}^{\boldsymbol{b}+1}}\prod_{l_1=1}^{d}L^{(a_{l_1})}\left(\frac{y_{l_1}-Y_{jl_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p}H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-W_{jl_2}}{\eta_{l_2}}\right), \tag{2.4.9}$$

so similar to Fan (1991b), one can consider substituting the variance $\mathrm{Var}(\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x}))$ with the sample variance of ξ_{nj}, $j=1,2,\ldots,n$. In fact, the conditions listed in Theorem 2.4.5 is already sufficient for doing so. However, the only unknown quantity in the variance function $V(y,x)$ defined in (2.4.8) is the joint density function of $f_{\boldsymbol{Y},\boldsymbol{W}}$, since it can be estimated by the classic kernel estimator which is consistent, so it provides another candidate for standardizing the variance in the asymptotic normal distribution.

The following result shows the asymptotic normality of the mixed kernel estimator when the super-smooth component is showing in the data. For convenience, we only consider the cases that the super-smooth measurement errors are symmetric. By doing so, we can avoid the discussion or real part and the imaginary part of the characteristic function of the super-smooth components.

Theorem 2.4.6. *Under conditions (S1)-(S4), we have*

$$\frac{\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})-E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})}{\sqrt{Var(\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}))}}\Longrightarrow N(0,1).$$

Furthermore, if

$$\sqrt{\frac{n\boldsymbol{h}_{\times}^{2\boldsymbol{a}+1}\boldsymbol{\eta}_{\times}^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}}{\boldsymbol{\rho}_{\times}^{\boldsymbol{\tau}}\prod_{l_3=1}^{q}\left[\exp\left(\frac{2-4\gamma_{l_3}b_{nl_3}}{\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}}}\right)\right]^2}}\left(\sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]}|\boldsymbol{h}|_{\times}^{\tilde{\boldsymbol{a}}}|\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}}|\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}}\right)(\boldsymbol{h}_{+}^{\alpha\mathbf{1}_d}+\boldsymbol{\eta}_{+}^{\alpha\mathbf{1}_p}+\boldsymbol{\rho}_{+}^{\alpha\mathbf{1}_q})\to 0,$$

where $b_{nl_3}=\rho_{l_3}^{\gamma_{l_3}/2(m+5)}$, $\boldsymbol{\tau}=\boldsymbol{\mu}_0[1+(m+4)/(2(m+5))]-2\boldsymbol{c}-1$, *then*

$$\frac{\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})-f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})}{\sqrt{Var(\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}))}}\Longrightarrow N(0,1).$$

Since the variance function $\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}))$ has no tractable form, to find a data-dependent replacement for the variance function $\mathrm{Var}(\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}))$ has some practical significance. Based on the assumptions in Theorem 2.4.6, we can show that $n^{-1}\sum_{j=1}^{n}\xi_{nj}^2$ can be served as such a replacement. The justification is similar to the Lemma 2.4 and Corollary 2.2 in Fan (1991b).

2.5 Simulation Study

To our best knowledge, the proposed estimator in this paper is the only procedure in literature to estimate the joint density function or its derivatives of multivariate random variables contaminated with different types of measurement errors. As a future research topic, we will explore the possibilities of modifying other classical density estimation procedures in error-free setups, such as the nearest neighbor, to account for the random variables measured with errors.

For illustration purpose, in this section we conduct a simulation study on estimating the joint density function of a 6-dimensional random vector $(Y_1,Y_2,X_1,X_2,Z_1,Z_2)^T$. Assume that $(Y_1,Y_2,X_1,X_2,Z_1,Z_2)^T$ is jointly normal with mean vector $\mathbf{0}$, and compound symmetric covariance matrix with equal variances 1 and covariance 0.25. Y_1 and Y_2 are directly observable, $W_1=X_1+U_1$, $W_2=X_2+U_2$ with U_1,U_2 being independent and distributed as the Laplace distribution with mean 0 and variance σ_u^2, $S_1 = Z_1 + V_1$, $S_2 = Z_2 + V_2$ with V_1, V_2 being independent and distributed as the normal distribution with mean 0 and variance σ_v^2. To be specific, we assume that

$$\begin{pmatrix} Y_1 \\ Y_2 \\ X_1 \\ X_2 \\ Z_1 \\ Z_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 1 \end{pmatrix} \right).$$

For the observed data (Y_1, Y_2), we choose the stand normal density function as the kernel function L; for the data (X_1, X_2) contaminated with the Laplace measurement error with mean 0 and variance σ_u^2, we choose H to be standard normal, the corresponding deconvolution kernel function H_n is defined as

$$H_n(x) = \phi(x) \left[1 - \frac{\sigma_u^2}{2h^2}(x^2 - 1) \right],$$

where $\phi(x)$ is the standard normal density function; for the data (Z_1, Z_2) contaminated with the normal measurement error with mean 0 and variance σ_v^2, we choose K to be the sinc kernel $K(x) = \sin(x)/(\pi x)$, the corresponding deconvolution kernel function K_n is defined as

$$K_n(x) = \sqrt{\frac{2}{\pi}} \int_0^1 \cos(tx) \exp\left(\frac{\sigma_v^2 t^2}{2h^2} \right) dt.$$

Based on a sample $(Y_{j1}, Y_{j2}, W_{j1}, W_{j2}, S_{j1}, S_{j2})^T, j = 1, 2, \ldots, n$, the mixed kernel density estimator $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ with equal bandwidth is defined as

$$\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) = \frac{1}{nh^6} \sum_{j=1}^{n} \left[L\left(\frac{y_1 - Y_{j1}}{h}\right) L\left(\frac{y_2 - Y_{j2}}{h}\right) H_n\left(\frac{x_1 - W_{j1}}{h}\right) H_n\left(\frac{x_2 - W_{j2}}{h}\right) K_n\left(\frac{z_1 - S_{j1}}{h}\right) K_n\left(\frac{z_2 - S_{j2}}{h}\right) \right].$$

The variances σ_u^2 and σ_v^2 are chosen to be 0.5^2. The sample sizes are chosen to be $n = 100$ to 1000, with a step of 100. At each sample size n, the estimators are evaluated at 100 randomly selected values from $[-2, 2]^6$, and the MSEs of these 100 estimated values are calculated. For comparison purpose, in addition to the mixed kernel estimator $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$, we also consider the mixed kernel estimators

$$\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{nh^4} \sum_{j=1}^{n} \left[L\left(\frac{y_1 - Y_{j1}}{h}\right) L\left(\frac{y_2 - Y_{j2}}{h}\right) H_n\left(\frac{x_1 - W_{j1}}{h}\right) H_n\left(\frac{x_2 - W_{j2}}{h}\right) \right]$$

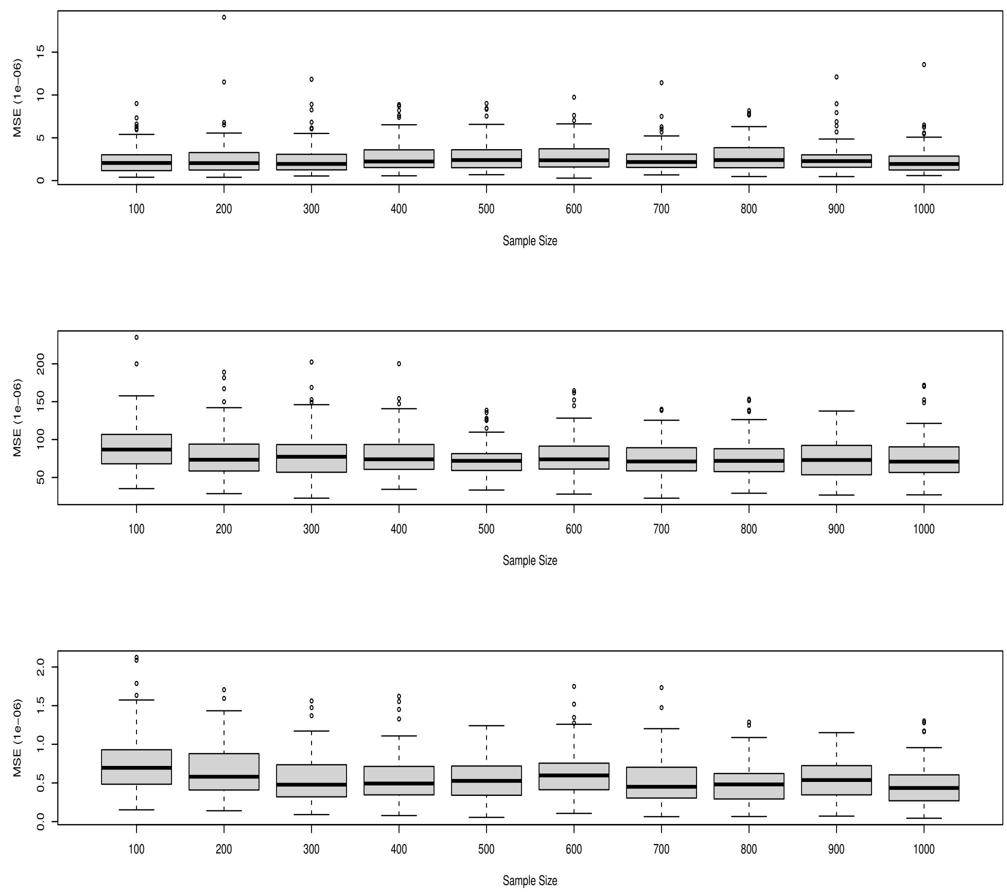
and

$$\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{z}) = \frac{1}{nh^4} \sum_{j=1}^{n} \left[L\left(\frac{y_1 - Y_{j1}}{h}\right) L\left(\frac{y_2 - Y_{j2}}{h}\right) K_n\left(\frac{z_1 - S_{j1}}{h}\right) K_n\left(\frac{z_2 - S_{j2}}{h}\right) \right].$$

When the super smooth component appears in the data, we choose $h = (\log n)^{-1/n}$. Note that the joint normal density function is a member of $\mathscr{C}_{0,0,B}$ for some B, we choose $h = n^{-1/12}$ $(2m + 2\alpha + 2[\boldsymbol{\beta}] + d + p = 0 + 0 + 8 + 2 + 2 = 12)$ when there is no super smooth component appears in the data. Each estimation procedure is repeated 100 times, and the boxplots of 100 MSEs at each sample size are shown in Figure 2.1.

From Figure 2.1, one can see that for $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ and $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{z})$, the decreasing trend, indicated by the median MSE values at various sample sizes, of the boxplots from $n = 100$ to $n = 1000$ is not clear. However, for $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x})$, the decreasing trend of the boxplots shows that its convergence to the true density function is much faster. We also did the simulation study for other σ_u^2 and σ_v^2 values, the resulting boxplots show similar patterns as in Figure 2.1, thus omitted for the sake of brevity.

Figure 2.1: Boxplots of 100 MSEs of the mixed kernel density estimators $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ (top), $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{z})$ (middle) and $\hat{f}_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x})$ (bottom), respectively. For each sample size, the estimator is evaluated at 100 randomly selected values from a hyper-rectangular domain with each coordinate in $[-2, 2]$, and its MSE is recorded. Each estimation procedure is then repeated 100 times, resulting in 100 MSEs.



2.6 Discussion

In this project, we established an optimal minimax convergence rates for nonparametric estimators of the multivariate density function, when the data are measured with different types of measurement errors with different smoothness within each type. We also showed that the mixed kernel and deconvolution kernel density estimators achieve the optimal convergence rate. However, in real applications, how to select a bandwidth for the estimators having a satisfactory finite sample performance still needs more extra efforts. It is noted that the theoretical optimal bandwidth is not just related to the sample size, but also relies on some constant factors which depend on some unknown quantities, even the density functions themselves. The direct plug-in method might be useful to partly solve the questions, but some data driven bandwidth selection procedures, similar to the least squares and biased cross validation methods discussed in Wand and Jones (1995), should be explored. Following the theoretical results we established for the density functions, another challenging research topic is to find a similar minimax convergence rate for the nonparametric estimators of the regression functions. Although we conjecture that the mixed kernel and deconvolution kernel based multivariate regression smoother most likely achieves the potential optimal convergence rate, its theoretical justification will be more difficult.

2.7 Proof of Main Results

The section contains all the technical proofs of the main results in Section 2.4.

Proof of Theorem 2.4.1: The expectation $E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$ equals

$$
\begin{aligned}
&\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}} E\left[\prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-Y_{l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-W_{l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3}-S_{l_3}}{\rho_{l_3}}\right)\right]\\
&=\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\int\!\!\int\!\!\int\!\!\int\!\!\int\!\!\int \prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-\tilde{x}_{l_2}-u_{l_2}}{\eta_{l_2}}\right)\cdot\\
&\qquad\prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3}-\tilde{z}_{l_3}-v_{l_3}}{\rho_{l_3}}\right) f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) f_{\boldsymbol{U}}(\boldsymbol{u}) f_{\boldsymbol{V}}(\boldsymbol{v})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}d\boldsymbol{u}d\boldsymbol{v}\\
&=\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\int\!\!\int\!\!\int\!\!\int\!\!\int\!\!\int \prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\cdot\\
&\qquad\prod_{l_2=1}^{p}\left[\frac{1}{2\pi}\int(-\mathbf{i}r_{l_2})^{b_{l_2}} e^{-\mathbf{i}r_{l_2}(x_{l_2}-\tilde{x}_{l_2}-u_{l_2})/\eta_{l_2}}\frac{\phi_H(r_{l_2})}{\phi_{U_{l_2}}(r_{l_2}/\eta_{l_2})}dr_{l_2}\right]\cdot\\
&\qquad\prod_{l_3=1}^{q}\left[\frac{1}{2\pi}\int(-\mathbf{i}t_{l_3})^{c_{l_3}} e^{-\mathbf{i}t_{l_3}(z_{l_3}-\tilde{z}_{l_3}-v_{l_3})/\rho_{l_3}}\frac{\phi_K(t_{l_3})}{\phi_{V_{l_3}}(t_{l_3}/\rho_{l_3})}dt_{l_3}\right]\cdot\\
&\qquad f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) f_{\boldsymbol{U}}(\boldsymbol{u}) f_{\boldsymbol{V}}(\boldsymbol{v})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}d\boldsymbol{u}d\boldsymbol{v}.
\end{aligned}
$$

By exchanging the order of integrations, we can rewrite $E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$ as

$$\begin{aligned}
&\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\iiint\prod_{l_1=1}^{d}L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\cdot\\
&\quad\prod_{l_2=1}^{p}\left[\frac{1}{2\pi}\int(-\mathbf{i}r_{l_2})^{b_{l_2}}e^{-\mathbf{i}r_{l_2}(x_{l_2}-\tilde{x}_{l_2})/\eta_{l_2}}\frac{\phi_H(r_{l_2})}{\phi_{U_{l_2}}(r_{l_2}/\eta_{l_2})}\left(\int e^{\mathbf{i}r_{l_2}u_{l_2}/\eta_{l_2}}f_{U_{l_2}}(u_{l_2})du_{l_2}\right)dr_{l_2}\right]\cdot\\
&\quad\prod_{l_3=1}^{q}\left[\frac{1}{2\pi}\int(-\mathbf{i}t_{l_3})^{c_{l_3}}e^{-\mathbf{i}t_{l_3}(z_{l_3}-\tilde{z}_{l_3})/\rho_{l_3}}\frac{\phi_K(t_{l_3})}{\phi_{V_{l_3}}(t_{l_3}/\rho_{l_3})}\left(\int e^{\mathbf{i}t_{l_3}v_{l_3}/\rho_{l_3}}f_{V_{l_3}}(v_{l_3})dv_{l_3}\right)dt_{l_3}\right]\cdot\\
&\quad f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}.
\end{aligned}$$

Note that $\int e^{\mathbf{i}r_{l_2}u_{l_2}/\eta_{l_2}}f_{U_{l_2}}(u_{l_2})du_{l_2}=\phi_{U_{l_2}}(r_{l_2}/\eta_{l_2})$, $\int e^{\mathbf{i}t_{l_3}v_{l_3}/\rho_{l_3}}f_{V_{l_3}}(v_{l_3})dv_{l_3}=\phi_{V_{l_3}}(t_{l_3}/\rho_{l_3})$, so we have

$$\begin{aligned}
&E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})=\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\cdot\\
&\quad\iiint\prod_{l_1=1}^{d}L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p}\left[\frac{1}{2\pi}\int(-\mathbf{i}r_{l_2})^{b_{l_2}}e^{-\mathbf{i}r_{l_2}(x_{l_2}-\tilde{x}_{l_2})/\eta_{l_2}}\phi_H(r_{l_2})dr_{l_2}\right]\cdot\\
&\quad\prod_{l_3=1}^{q}\left[\frac{1}{2\pi}\int(-\mathbf{i}t_{l_3})^{c_{l_3}}e^{-\mathbf{i}t_{l_3}(z_{l_3}-\tilde{z}_{l_3})/\rho_{l_3}}\phi_K(t_{l_3})dt_{l_3}\right]f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}\\
&=\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\iiint\prod_{l_1=1}^{d}L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\cdot\\
&\quad\prod_{l_2=1}^{p}\eta_{l_2}^{b_{l_2}}\frac{\partial^{b_{l_2}}}{\partial x_{l_2}^{b_{l_2}}}\left[\frac{1}{2\pi}\int e^{-\mathbf{i}r_{l_2}(x_{l_2}-\tilde{x}_{l_2})/\eta_{l_2}}\phi_H(r_{l_2})dr_{l_2}\right]\cdot\\
&\quad\prod_{l_3=1}^{q}\rho_{l_3}^{c_{l_3}}\frac{\partial^{c_{l_3}}}{\partial z_{l_3}^{c_{l_3}}}\left[\frac{1}{2\pi}\int e^{-\mathbf{i}t_{l_3}(z_{l_3}-\tilde{z}_{l_3})/\rho_{l_3}}\phi_K(t_{l_3})dt_{l_3}\right]f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}} \iiint \prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right) \prod_{l_2=1}^{p} \eta_{l_2}^{b_{l_2}} \frac{\partial^{b_{l_2}}}{\partial x_{l_2}^{b_{l_2}}} H\left(\frac{x_{l_2}-\tilde{x}_{l_2}}{\eta_{l_2}}\right) \cdot \\
&\quad \prod_{l_3=1}^{q} \eta_{l_3}^{c_{l_3}} \frac{\partial^{c_{l_3}}}{\partial z_{l_3}^{c_{l_3}}} K\left(\frac{z_{l_3}-\tilde{z}_{l_3}}{\rho_{l_3}}\right) f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}} \\
&= \frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}} \iiint \prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right) \prod_{l_2=1}^{p} H^{(b_{l_2})}\left(\frac{x_{l_2}-\tilde{x}_{l_2}}{\eta_{l_2}}\right) \cdot \\
&\quad \prod_{l_3=1}^{q} K^{(c_{l_3})}\left(\frac{z_{l_3}-\tilde{z}_{l_3}}{\rho_{l_3}}\right) f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}}. \qquad (2.7.1)
\end{aligned}
$$

To proceed, let us compute the following integration:

$$
\int \prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right) f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) d\tilde{\boldsymbol{y}}.
$$

In fact, by changing integration orders, the above integral can be written as

$$
\begin{aligned}
&\int L^{(a_1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right) \cdots L^{(a_d)}\left(\frac{y_d-\tilde{y}_d}{h_d}\right) f_{\boldsymbol{D}}(\tilde{y}_1,\cdots,\tilde{y}_d,\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) d\tilde{\boldsymbol{y}} \qquad (2.7.2) \\
&= \int \left[\int L^{(a_1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right) f_{\boldsymbol{D}}(\tilde{y}_1,\cdots,\tilde{y}_d,\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) d\tilde{y}_1\right] \prod_{l_1=2}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right) d\tilde{y}_2\cdots d\tilde{y}_d.
\end{aligned}
$$

Integration by parts, from condition (C1), we have

$$
\begin{aligned}
&\int L^{(a_1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right) f_{\boldsymbol{D}}(\tilde{y}_1,\cdots,\tilde{y}_d,\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) d\tilde{y}_1 = -h_1 \int f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}}) dL^{(a_1-1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right) \\
&= -h_1 L^{(a_1-1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right) f_{\boldsymbol{D}}(\tilde{y}_1,\cdots,\tilde{y}_d,\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})\Big|_{-\infty}^{\infty} + h_1 \int \frac{\partial f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{y}_1} L^{(a_1-1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right) d\tilde{y}_1 \\
&= h_1^{a_1} \int \frac{\partial^{a_1} f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{y}_1^{a_1}} L\left(\frac{y_1-\tilde{y}_1}{h_1}\right) d\tilde{y}_1.
\end{aligned}
$$

Plugging the above result into (3.3.1) and repeat the above procedure for $\tilde{y}_2$, $\tilde{y}_3$, …, and $\tilde{y}_d$, we get

$$
\begin{aligned}
&\int L^{(a_1)}\left(\frac{y_1-\tilde{y}_1}{h_1}\right)\cdots L^{(a_d)}\left(\frac{y_d-\tilde{y}_d}{h_d}\right) f_{\boldsymbol{D}}(\tilde{y}_1,\cdots,\tilde{y}_d,\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})d\tilde{\boldsymbol{y}}\\
&=h_1^{a_1}\int \frac{\partial^{a_1} f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{y}_1^{a_1}} L\left(\frac{y_1-\tilde{y}_1}{h_1}\right)\prod_{l_1=2}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)d\tilde{\boldsymbol{y}}\\
&=h_1^{a_1}h_2^{a_2}\int \frac{\partial^{a_1+a_2} f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{y}_1^{a_1}\partial \tilde{y}_2^{a_2}} L\left(\frac{y_1-\tilde{y}_1}{h_1}\right)L\left(\frac{y_2-\tilde{y}_2}{h_2}\right)\prod_{l_1=3}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)d\tilde{\boldsymbol{y}}\\
&=\boldsymbol{h}_{\times}^{\boldsymbol{a}}\int \frac{\partial^{[\boldsymbol{a}]} f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{\boldsymbol{y}}^{\boldsymbol{a}}}\prod_{l_1=1}^{d} L\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)d\tilde{\boldsymbol{y}},
\end{aligned}
$$

where $\partial\tilde{\boldsymbol{y}}^{\boldsymbol{a}} = \partial\tilde{y}_1^{a_1}\partial\tilde{y}_2^{a_2}\cdots\partial\tilde{y}_d^{a_d}$. Plugging the above result into (2.7.1), repeat the above procedures for the derivatives of H and K, then changing variables, we can get

$$
\begin{aligned}
&E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\\
&=\frac{1}{\boldsymbol{h}_{\times}^{\boldsymbol{J}_d}\boldsymbol{\eta}_{\times}^{\boldsymbol{b}+1}\boldsymbol{\rho}_{\times}^{\boldsymbol{c}+1}}\iiint\prod_{l_1=1}^{d} L\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H^{(b_{l_2})}\left(\frac{x_{l_2}-\tilde{x}_{l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K^{(c_{l_3})}\left(\frac{z_{l_3}-\tilde{z}_{l_3}}{\rho_{l_3}}\right)\cdot\\
&\quad\frac{\partial^{[\boldsymbol{a}]} f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{\boldsymbol{y}}^{\boldsymbol{a}}}d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}\\
&=\frac{1}{\boldsymbol{h}_{\times}^{\boldsymbol{J}_d}\boldsymbol{\eta}_{\times}^{\boldsymbol{J}_p}\boldsymbol{\rho}_{\times}^{\boldsymbol{J}_q}}\cdot\\
&\quad\iiint\prod_{l_1=1}^{d} L\left(\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H\left(\frac{x_{l_2}-\tilde{x}_{l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K\left(\frac{z_{l_3}-\tilde{z}_{l_3}}{\rho_{l_3}}\right)\frac{\partial^{[\boldsymbol{k}]} f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})}{\partial \tilde{\boldsymbol{y}}^{\boldsymbol{a}}\partial \tilde{\boldsymbol{x}}^{\boldsymbol{b}}\partial \tilde{\boldsymbol{z}}^{\boldsymbol{c}}}d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}\\
&=\iiint\prod_{l_1=1}^{d} L(\tilde{y}_{l_1})\prod_{l_2=1}^{p} H(\tilde{x}_{l_2})\prod_{l_3=1}^{q} K(\tilde{z}_{l_3})\frac{\partial^{[\boldsymbol{k}]} f_{\boldsymbol{D}}(\boldsymbol{y}^*,\boldsymbol{x}^*,\boldsymbol{z}^*)}{\partial \boldsymbol{y}^{*\boldsymbol{a}}\partial \boldsymbol{x}^{*\boldsymbol{b}}\partial \boldsymbol{z}^{*\boldsymbol{c}}}\Bigg|_{\substack{\boldsymbol{y}^*=\boldsymbol{y}-\tilde{\boldsymbol{y}}\boldsymbol{h}\\ \boldsymbol{x}^*=\boldsymbol{x}-\tilde{\boldsymbol{x}}\boldsymbol{\eta}\\ \boldsymbol{z}^*=\boldsymbol{z}-\tilde{\boldsymbol{z}}\boldsymbol{\rho}}}d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}.
\end{aligned}
\tag{2.7.3}
$$

Denote $\tilde{\boldsymbol{k}} = (\tilde{\boldsymbol{a}}^T, \tilde{\boldsymbol{b}}^T, \tilde{\boldsymbol{c}}^T)^T$. By Taylor expansion up to the m-th order, we have

$$
\begin{aligned}
&\frac{\partial^{[\boldsymbol{k}]} f_{\boldsymbol{D}}(\boldsymbol{y}^*, \boldsymbol{x}^*, \boldsymbol{z}^*)}{\partial \boldsymbol{y}^{*\boldsymbol{a}} \partial \boldsymbol{x}^{*\boldsymbol{b}} \partial \boldsymbol{z}^{*\boldsymbol{c}}}\Bigg|_{\substack{\boldsymbol{y}^* = \boldsymbol{y} - \tilde{\boldsymbol{y}}\boldsymbol{h} \\ \boldsymbol{x}^* = \boldsymbol{x} - \tilde{\boldsymbol{x}}\boldsymbol{\eta} \\ \boldsymbol{z}^* = \boldsymbol{z} - \tilde{\boldsymbol{z}}\boldsymbol{\rho}}} \\
= &\frac{\partial^{[\boldsymbol{k}]} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\boldsymbol{a}} \partial \boldsymbol{x}^{\boldsymbol{b}} \partial \boldsymbol{z}^{\boldsymbol{c}}} + \sum_{[\tilde{\boldsymbol{k}}]=1}^{m-[\boldsymbol{k}]-1} \frac{1}{\tilde{\boldsymbol{a}}! \tilde{\boldsymbol{b}}! \tilde{\boldsymbol{c}}!} \frac{\partial^{[\boldsymbol{k}]+[\tilde{\boldsymbol{k}}]} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}+\boldsymbol{a}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}+\boldsymbol{b}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}+\boldsymbol{c}}} (-\tilde{\boldsymbol{y}}\boldsymbol{h})_{\times}^{\tilde{\boldsymbol{a}}} (-\tilde{\boldsymbol{x}}\boldsymbol{\eta})_{\times}^{\tilde{\boldsymbol{b}}} (-\tilde{\boldsymbol{z}}\boldsymbol{\rho})_{\times}^{\tilde{\boldsymbol{c}}} \\
&+ \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \frac{1}{\tilde{\boldsymbol{a}}! \tilde{\boldsymbol{b}}! \tilde{\boldsymbol{c}}!} \frac{\partial^{m} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}+\boldsymbol{a}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}+\boldsymbol{b}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}+\boldsymbol{c}}}\Bigg|_{\substack{\boldsymbol{y} = \tilde{\boldsymbol{y}}^* \\ \boldsymbol{x} = \tilde{\boldsymbol{x}}^* \\ \boldsymbol{z} = \tilde{\boldsymbol{z}}^*}} (-\tilde{\boldsymbol{y}}\boldsymbol{h})_{\times}^{\tilde{\boldsymbol{a}}} (-\tilde{\boldsymbol{x}}\boldsymbol{\eta})_{\times}^{\tilde{\boldsymbol{b}}} (-\tilde{\boldsymbol{z}}\boldsymbol{\rho})_{\times}^{\tilde{\boldsymbol{c}}}.
\end{aligned}
$$

From condition (C2), $E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ can be written as

$$
\begin{aligned}
&\frac{\partial^{[\boldsymbol{k}]} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\boldsymbol{a}} \partial \boldsymbol{x}^{\boldsymbol{b}} \partial \boldsymbol{z}^{\boldsymbol{c}}} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \frac{1}{\tilde{\boldsymbol{a}}! \tilde{\boldsymbol{b}}! \tilde{\boldsymbol{c}}!} \iiint \prod_{l_1=1}^{d} L(\tilde{y}_{l_1}) \prod_{l_2=1}^{p} H(\tilde{x}_{l_2}) \prod_{l_3=1}^{q} K(\tilde{z}_{l_3}) \cdot \\
&\qquad \frac{\partial^{m} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}+\boldsymbol{a}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}+\boldsymbol{b}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}+\boldsymbol{c}}}\Bigg|_{\substack{\boldsymbol{y} = \tilde{\boldsymbol{y}}^* \\ \boldsymbol{x} = \tilde{\boldsymbol{x}}^* \\ \boldsymbol{z} = \tilde{\boldsymbol{z}}^*}} (-\tilde{\boldsymbol{y}}\boldsymbol{h})_{\times}^{\tilde{\boldsymbol{a}}} (-\tilde{\boldsymbol{x}}\boldsymbol{\eta})_{\times}^{\tilde{\boldsymbol{b}}} (-\tilde{\boldsymbol{z}}\boldsymbol{\rho})_{\times}^{\tilde{\boldsymbol{c}}} d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}}.
\end{aligned}
$$

Since the density function $f_{\boldsymbol{D}}$ is a member of the class $\mathscr{C}_{\alpha,m,B}$, we can get

$$
\begin{aligned}
&|E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})| \\
= &\Bigg| \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \frac{1}{\tilde{\boldsymbol{a}}! \tilde{\boldsymbol{b}}! \tilde{\boldsymbol{c}}!} \iiint (-\tilde{\boldsymbol{y}}\boldsymbol{h})_{\times}^{\tilde{\boldsymbol{a}}} (-\tilde{\boldsymbol{x}}\boldsymbol{\eta})_{\times}^{\tilde{\boldsymbol{b}}} (-\tilde{\boldsymbol{z}}\boldsymbol{\rho})_{\times}^{\tilde{\boldsymbol{c}}} \prod_{l_1=1}^{d} L(\tilde{y}_{l_1}) \prod_{l_2=1}^{p} H(\tilde{x}_{l_2}) \prod_{l_3=1}^{q} K(\tilde{z}_{l_3}) \cdot \\
&\qquad \left[\frac{\partial^{m} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}+\boldsymbol{a}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}+\boldsymbol{b}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}+\boldsymbol{c}}}\Bigg|_{\substack{\boldsymbol{y} = \tilde{\boldsymbol{y}}^* \\ \boldsymbol{x} = \tilde{\boldsymbol{x}}^* \\ \boldsymbol{z} = \tilde{\boldsymbol{z}}^*}} - \frac{\partial^{m} f_{\boldsymbol{D}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})}{\partial \boldsymbol{y}^{\tilde{\boldsymbol{a}}+\boldsymbol{a}} \partial \boldsymbol{x}^{\tilde{\boldsymbol{b}}+\boldsymbol{b}} \partial \boldsymbol{z}^{\tilde{\boldsymbol{c}}+\boldsymbol{c}}} \right] d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}} \Bigg|,
\end{aligned}
$$

which is bounded above by

$$B \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \frac{1}{\tilde{\mathbf{a}}!\tilde{\boldsymbol{b}}!\tilde{\boldsymbol{c}}!} \iiint |\tilde{\boldsymbol{y}}\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}} |\tilde{\boldsymbol{x}}\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}} |\tilde{\boldsymbol{z}}\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}} | \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| \cdot$$
$$(|\tilde{\boldsymbol{y}}\boldsymbol{h}|_{+}^{\alpha} + |\tilde{\boldsymbol{x}}\boldsymbol{\eta}|_{+}^{\alpha} + |\tilde{\boldsymbol{z}}\boldsymbol{\rho}|_{+}^{\alpha}) d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}}.$$

Note that $|\tilde{\boldsymbol{y}}\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}} \cdot [\tilde{\boldsymbol{y}}\boldsymbol{h}]_{+}^{\alpha} = \sum_{j=1}^{d} |\tilde{\boldsymbol{y}}|_{\times}^{\tilde{\mathbf{a}}+\alpha\mathbf{1}_{dj}} |h|_{\times}^{\tilde{\mathbf{a}}+\alpha\mathbf{1}_{dj}}$, we have $|E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})|$ to be bounded above by

$$B \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \frac{1}{\tilde{\mathbf{a}}!\tilde{\boldsymbol{b}}!\tilde{\boldsymbol{c}}!} \iiint \prod_{l_1=1}^{d} |L(\tilde{y}_{l_1})| \prod_{l_2=1}^{p} |H(\tilde{x}_{l_2})| \prod_{l_3=1}^{q} |K(\tilde{z}_{l_3})| \cdot$$
$$\Bigg[|\tilde{\boldsymbol{x}}\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}} |\tilde{\boldsymbol{z}}\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}} \sum_{j_1=1}^{d} |\tilde{\boldsymbol{y}}|_{\times}^{\tilde{\mathbf{a}}+\alpha\mathbf{1}_{dj_1}} |\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}+\alpha\mathbf{1}_{dj_1}} + |\tilde{\boldsymbol{y}}\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}} |\tilde{\boldsymbol{z}}\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}} \sum_{j_2=1}^{p} |\tilde{\boldsymbol{x}}|_{\times}^{\tilde{\boldsymbol{b}}+\alpha\mathbf{1}_{pj_2}} |\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}+\alpha\mathbf{1}_{pj_2}}$$
$$+ |\tilde{\boldsymbol{y}}\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}} |\tilde{\boldsymbol{x}}\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}} \sum_{j_3=1}^{q} |\tilde{\boldsymbol{z}}|_{\times}^{\tilde{\boldsymbol{c}}+\alpha\mathbf{1}_{qj_3}} |\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}+\alpha\mathbf{1}_{qj_3}} \Bigg] d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{z}}$$
$$= \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_1=1}^{d} B_{j_1,\tilde{\boldsymbol{k}},1} |\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}+\alpha\mathbf{1}_{dj_1}} |\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}} |\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}} + \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_2=1}^{p} B_{j_2,\tilde{\boldsymbol{k}},2} |\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}} |\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}+\alpha\mathbf{1}_{pj_2}} |\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}}$$
$$+ \sum_{[\tilde{\boldsymbol{k}}]=m-[\boldsymbol{k}]} \sum_{j_3=1}^{q} B_{j_3,\tilde{\boldsymbol{k}},3} |\boldsymbol{h}|_{\times}^{\tilde{\mathbf{a}}} |\boldsymbol{\eta}|_{\times}^{\tilde{\boldsymbol{b}}} |\boldsymbol{\rho}|_{\times}^{\tilde{\boldsymbol{c}}+\alpha\mathbf{1}_{qj_3}}. \tag{2.7.4}$$

Next, let us compute the variance of $\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$. Since $(\boldsymbol{Y}_j, \boldsymbol{W}_j, \boldsymbol{S}_j)$, $j = 1, 2, \ldots, n$ are independent and identically distributed, so we have

$$E[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})]^2 = \frac{1}{n\boldsymbol{h}_{\times}^{2(\boldsymbol{a}+1)} \boldsymbol{\eta}_{\times}^{2(\boldsymbol{b}+1)} \boldsymbol{\rho}_{\times}^{2(\boldsymbol{c}+1)}} \cdot$$
$$\left(E\left[\prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1} - Y_{1l_1}}{h_{l_1}}\right) \prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2} - W_{1l_2}}{\eta_{l_2}}\right) \prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3} - S_{1l_3}}{\rho_{l_3}}\right) \right]^2 \right.$$

$$
-\left[E\prod_{l_1=1}^{d}L^{(a_{l_1})}\left(\frac{y_{l_1}-Y_{1l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p}H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-W_{1l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q}K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3}-S_{1l_3}}{\rho_{l_3}}\right)\right]^2\Bigg).
$$

From the derivation of bias, we already know that the squared expectation is the order of $O(1)$.

Note that

$$
\begin{aligned}
&E\left|\prod_{l_1=1}^{d}L^{(a_{l_1})}\left(\frac{y_{l_1}-Y_{1l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p}H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-W_{1l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q}K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3}-S_{1l_3}}{\rho_{l_3}}\right)\right|^2\\
=&E\Bigg|\frac{1}{(2\pi)^{d+p+q}}\prod_{l_1=1}^{d}\int(-\mathbf{i}r_{l_1})^{a_{l_1}}\exp\left[-\mathbf{i}r_{l_1}\frac{y_{l_1}-Y_{l_1}}{h_{l_1}}\right]\phi_L(r_{l_1})dr_{l_1}\cdot\\
&\prod_{l_2=1}^{p}\int(-\mathbf{i}t_{l_2})^{b_{l_2}}\exp\left[-\mathbf{i}t_{l_2}\frac{x_{l_2}-W_{l_2}}{\eta_{l_2}}\right]\phi_H(t_{l_2})\phi_U^{-1}(t_{l_2}/\eta_{l_2})dt_{l_2}\cdot\\
&\prod_{l_3=1}^{q}\int(-\mathbf{i}v_{l_3})^{c_{l_3}}\exp\left[-\mathbf{i}v_{l_3}\frac{z_{l_3}-S_{l_3}}{\rho_{l_3}}\right]\phi_K(v_{l_3})\phi_V^{-1}(v_{l_3}/\rho_{l_3})dv_{l_3}\Bigg|^2.
\end{aligned}
$$

To proceed, we may treat the super-smooth component differently by following Fan (1991a)'s argument. Note that the above expectation is bounded above by

$$
\begin{aligned}
&\frac{1}{(2\pi)^{2(d+p+q)}}E\Bigg|\prod_{l_1=1}^{d}\int(-\mathbf{i}r_{l_1})^{a_{l_1}}\exp\left[-\mathbf{i}r_{l_1}\frac{y_{l_1}-Y_{l_1}}{h_{l_1}}\right]\phi_L(r_{l_1})dr_{l_1}\cdot\\
&\prod_{l_2=1}^{p}\int(-\mathbf{i}t_{l_2})^{b_{l_2}}\exp\left[-\mathbf{i}t_{l_2}\frac{x_{l_2}-W_{l_2}}{\eta_{l_2}}\right]\phi_H(t_{l_2})\phi_U^{-1}(t_{l_2}/\eta_{l_2})dt_{l_2}\Bigg|^2\cdot\\
&\prod_{l_3=1}^{q}\left|\int_{-1}^{1}|\phi_K(v_{l_3})\phi_V^{-1}(v_{l_3}/\rho_{l_3})|dv_{l_3}\right|^2
\end{aligned}
$$

If we assume that the density function of $\boldsymbol{Y}, \boldsymbol{X}$, $f_{\boldsymbol{D}}(\boldsymbol{Y}, \boldsymbol{X})$, is bounded, then

$$
\begin{aligned}
&\frac{1}{(2\pi)^{2(d+p)}} E\Bigg| \prod_{l_1=1}^{d} \int (-\mathbf{i} r_{l_1})^{a_{l_1}} \exp\left[-\mathbf{i} r_{l_1} \frac{y_{l_1} - Y_{l_1}}{h_{l_1}} \right] \phi_L(r_{l_1}) dr_{l_1} \cdot \\
&\qquad \prod_{l_2=1}^{p} \int (-\mathbf{i} t_{l_2})^{b_{l_2}} \exp\left[-\mathbf{i} t_{l_2} \frac{x_{l_2} - W_{l_2}}{\eta_{l_2}} \right] \phi_H(t_{l_2}) \phi_U^{-1}(t_{l_2}/\eta_{l_2}) dt_{l_2} \Bigg|^2 \\
&= \frac{1}{(2\pi)^{2(d+p)}} \iiint \Bigg| \prod_{l_1=1}^{d} \int (-\mathbf{i} r_{l_1})^{a_{l_1}} \exp\left[-\mathbf{i} r_{l_1} \frac{y_{l_1} - \tilde{y}_{l_1}}{h_{l_1}} \right] \phi_L(r_{l_1}) dr_{l_1} \cdot \\
&\qquad \prod_{l_2=1}^{p} \int (-\mathbf{i} t_{l_2})^{b_{l_2}} \exp\left[-\mathbf{i} t_{l_2} \frac{x_{l_2} - \tilde{x}_{l_2} - \tilde{u}_{l_2}}{\eta_{l_2}} \right] \phi_H(t_{l_2}) \phi_U^{-1}(t_{l_2}/\eta_{l_2}) dt_{l_2} \Bigg|^2 \cdot \\
&\qquad f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{x}}) f_{\boldsymbol{U}}(\tilde{\boldsymbol{u}}) d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{u}} \\
&= \frac{\boldsymbol{h}_\times \boldsymbol{\eta}_\times}{(2\pi)^{2(d+p)}} \iiint \Bigg| \prod_{l_1=1}^{d} \int (-\mathbf{i} r_{l_1})^{a_{l_1}} \exp\left[-\mathbf{i} r_{l_1} \tilde{y}_{l_1} \right] \phi_L(r_{l_1}) dr_{l_1} \cdot \\
&\qquad \prod_{l_2=1}^{p} \int (-\mathbf{i} t_{l_2})^{b_{l_2}} \exp\left[-\mathbf{i} t_{l_2} \tilde{x}_{l_2} \right] \phi_H(t_{l_2}) \phi_U^{-1}(t_{l_2}/\eta_{l_2}) dt_{l_2} \Bigg|^2 \cdot \\
&\qquad f_{\boldsymbol{D}}(\boldsymbol{y} - \boldsymbol{h}\tilde{\boldsymbol{y}}, \boldsymbol{x} - \tilde{\boldsymbol{u}} - \boldsymbol{\eta}\tilde{\boldsymbol{x}}) f_{\boldsymbol{U}}(\tilde{\boldsymbol{u}}) d\tilde{\boldsymbol{y}} d\tilde{\boldsymbol{x}} d\tilde{\boldsymbol{u}},
\end{aligned}
$$

which is bounded above by $c(2\pi)^{-(d+p)} \boldsymbol{h}_\times \boldsymbol{\eta}_\times \prod_{l_2=1}^{p} \int |t_{l_2}|^{2b_{l_2}} \phi_H^2(t_{l_2}) \phi_U^{-2}(t_{l_2}/\eta_{l_2}) dt_{l_2}$, where c is a positive constant. The last inequality is from the boundedness of $f_{\boldsymbol{D}}$ and the Parseval identity. Therefore, from Fan (1991a),

$$
\int |t_{l_2}|^{2b_{l_2}} \phi_H^2(t_{l_2}) \phi_U^{-2}(t_{l_2}/\eta_{l_2}) dt_{l_2} = O(\eta_{l_2}^{-2\beta_{l_2}}), \tag{2.7.5}
$$

and

$$
\int |v_{l_3}|^{2c_{l_3}} \phi_K^2(v_{l_3}) \phi_V^{-2}(v_{l_3}/\rho_{l_3}) dv_{l_3} = O\left(\rho_{l_3}^{2\bar{b}_{l_3}} \exp(2\rho_{l_3}^{-\gamma_{l_3}} \xi_{l_3}^{-1}) \right), \tag{2.7.6}
$$

where $\bar{b}_{l_3} = 0$ if $\mu_{l_3 0} \geq 0$ and $\mu_{l_3 0}$ if $\mu_{l_3 0} < 0$, we have

$$
\begin{aligned}
&E\left| \prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1} - Y_{1l_1}}{h_{l_1}} \right) \prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2} - W_{1l_2}}{\eta_{l_2}} \right) \prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3} - S_{1l_3}}{\rho_{l_3}} \right) \right|^2 \\
&= O(\boldsymbol{h}_\times \boldsymbol{\eta}_\times^{1-2\boldsymbol{\beta}} \boldsymbol{\rho}_\times^{2\bar{\boldsymbol{b}}} \exp((2\boldsymbol{\rho}^{-\gamma} \boldsymbol{\xi}^{-1})_+)).
\end{aligned} \tag{2.7.7}
$$

If we further assume that $f_{\boldsymbol{D}}$ is bounded, we obtain

$$E\left|\prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-Y_{1l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-W_{1l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3}-S_{1l_3}}{\rho_{l_3}}\right)\right|^2$$

$$\begin{aligned}
&= \iiiint\!\!\int \left[\frac{1}{(2\pi)^{d+p+q}}\prod_{l_1=1}^{d}\int(-\mathbf{i}r_{l_1})^{a_{l_1}}\exp\left[-\mathbf{i}r_{l_1}\frac{y_{l_1}-\tilde{y}_{l_1}}{h_{l_1}}\right]\phi_L(r_{l_1})dr_{l_1}\cdot\right.\\
&\qquad \prod_{l_2=1}^{p}\int(-\mathbf{i}t_{l_2})^{b_{l_2}}\exp\left[-\mathbf{i}t_{l_2}\frac{x_{l_2}-\tilde{x}_{l_2}-\tilde{u}_{l_2}}{\eta_{l_2}}\right]\phi_H(t_{l_2})\phi_U^{-1}(t_{l_2}/\eta_{l_2})dt_{l_2}\cdot\\
&\qquad \left.\prod_{l_3=1}^{q}\int(-\mathbf{i}v_{l_3})^{c_{l_3}}\exp\left[-\mathbf{i}v_{l_3}\frac{z_{l_3}-\tilde{z}_{l_3}-\tilde{v}_{l_3}}{\rho_{l_3}}\right]\phi_K(v_{l_3})\phi_V^{-1}(v_{l_3}/\rho_{l_3})dv_{l_3}\right]^2\cdot\\
&\qquad f_{\boldsymbol{D}}(\tilde{\boldsymbol{y}},\tilde{\boldsymbol{x}},\tilde{\boldsymbol{z}})f_U(\tilde{\boldsymbol{u}})f_V(\tilde{\boldsymbol{v}})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}d\tilde{\boldsymbol{u}}d\tilde{\boldsymbol{v}}\\
&= \frac{(\boldsymbol{h\eta\rho})_{\times}}{(2\pi)^{2(d+p+q)}}\iiiint\!\!\int\left[\prod_{l_1=1}^{d}\int(-\mathbf{i}r_{l_1})^{a_{l_1}}\exp\left[-\mathbf{i}r_{l_1}\tilde{y}_{l_1}\right]\phi_L(r_{l_1})dr_{l_1}\right]^2\cdot\\
&\qquad \left[\prod_{l_2=1}^{p}\int(-\mathbf{i}t_{l_2})^{b_{l_2}}\exp\left[-\mathbf{i}t_{l_2}\tilde{x}_{l_2}\right]\phi_H(t_{l_2})\phi_U^{-1}(t_{l_2}/\eta_{l_2})dt_{l_2}\right]^2\cdot\\
&\qquad \left[\prod_{l_3=1}^{q}\int(-\mathbf{i}v_{l_3})^{c_{l_3}}\exp\left[-\mathbf{i}v_{l_3}\tilde{z}_{l_3}\right]\phi_K(v_{l_3})\phi_V^{-1}(v_{l_3}/\rho_{l_3})dv_{l_3}\right]^2\cdot\\
&\qquad f_{\boldsymbol{D}}(\boldsymbol{y}-\tilde{\boldsymbol{y}}\boldsymbol{h},\boldsymbol{x}-\tilde{\boldsymbol{u}}-\tilde{\boldsymbol{x}}\boldsymbol{\eta},\boldsymbol{z}-\tilde{\boldsymbol{v}}-\tilde{\boldsymbol{z}}\boldsymbol{\rho})f_U(\tilde{\boldsymbol{u}})f_V(\tilde{\boldsymbol{v}})d\tilde{\boldsymbol{y}}d\tilde{\boldsymbol{x}}d\tilde{\boldsymbol{z}}d\tilde{\boldsymbol{u}}d\tilde{\boldsymbol{v}},
\end{aligned}$$

which is bounded above by

$$\begin{aligned}
&\frac{(\boldsymbol{h\eta\rho})_{\times}}{(2\pi)^{2(d+p+q)}}\prod_{l_1=1}^{d}\int|r_{l_1}|^{2a_{l_1}}\phi_L^2(r_{l_1})dr_{l_1}\cdot\prod_{l_2=1}^{p}\int|t_{l_2}|^{2b_{l_2}}\phi_H^2(t_{l_2})\phi_U^{-2}(t_{l_2}/\eta_{l_2})dt_{l_2}\cdot\\
&\qquad \prod_{l_3=1}^{q}\int|v_{l_3}|^{2c_{l_3}}\phi_K^2(v_{l_3})\phi_V^{-2}(v_{l_3}/\rho_{l_3})dv_{l_3}.
\end{aligned}$$

Again using (2.7.5) and (2.7.6), we have

$$E\left[\prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1}-Y_{1l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2}-W_{1l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3}-S_{1l_3}}{\rho_{l_3}}\right)\right]^2 \tag{2.7.8}$$
$$= O\left((\boldsymbol{h}\boldsymbol{\eta}\boldsymbol{\rho})_\times \boldsymbol{\eta}_\times^{-2\boldsymbol{\beta}}\boldsymbol{\rho}^{2\bar{\boldsymbol{b}}}\exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}}\boldsymbol{\xi}^{-1})_+)\right) = O\left(\boldsymbol{h}_\times \boldsymbol{\eta}_\times^{1-2\boldsymbol{\beta}}\boldsymbol{\rho}_\times^{1+2\bar{\boldsymbol{b}}}\exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}}\boldsymbol{\xi}^{-1})_+)\right).$$

Hence if using (2.7.7)

$$\begin{aligned}
&E[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})]^2 \\
&= \frac{1}{n\boldsymbol{h}_\times^{2(\boldsymbol{a}+1)}\boldsymbol{\eta}_\times^{2(\boldsymbol{b}+1)}\boldsymbol{\rho}_\times^{2(\boldsymbol{c}+1)}}\cdot O(\boldsymbol{h}_\times \boldsymbol{\eta}_\times^{1-2\boldsymbol{\beta}}\boldsymbol{\rho}_\times^{2\bar{\boldsymbol{b}}}\exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}}\boldsymbol{\xi}^{-1})_+)) \\
&= O\left(\frac{1}{n\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}\boldsymbol{\rho}_\times^{2\boldsymbol{c}-2\bar{\boldsymbol{b}}+2}}\exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}}\boldsymbol{\xi}^{-1})_+)\right).
\end{aligned} \tag{2.7.9}$$

If using (2.7.8),

$$\begin{aligned}
&E[\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - E\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})]^2 \\
&= \frac{1}{n\boldsymbol{h}_\times^{2(\boldsymbol{a}+1)}\boldsymbol{\eta}_\times^{2(\boldsymbol{b}+1)}\boldsymbol{\rho}_\times^{2(\boldsymbol{c}+1)}}\cdot O\left(\boldsymbol{h}_\times \boldsymbol{\eta}_\times^{1-2\boldsymbol{\beta}}\boldsymbol{\rho}_\times^{1+2\bar{\boldsymbol{b}}}\exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}}\boldsymbol{\xi}^{-1})_+)\right) \\
&= O\left(\frac{1}{n\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}\boldsymbol{\rho}_\times^{2\boldsymbol{c}-2\bar{\boldsymbol{b}}+1}}\exp((2\boldsymbol{\rho}^{-\boldsymbol{\gamma}}\boldsymbol{\xi}^{-1})_+)\right).
\end{aligned} \tag{2.7.10}$$

Since $\boldsymbol{\rho} \to 0$ as $n \to \infty$, so (2.7.8) also has the order of (2.7.7). Note that (3.3.2), (2.7.9) and (2.7.10) are uniform in $\mathscr{C}_{m,\alpha,B}$, finally we obtain (2.4.2).

Proof of Theorem 2.4.4: Let $\lambda = m + \alpha$, and define

$$f_0(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) = \prod_{l_1=1}^{d}\frac{C_{1l_1}}{(1+y_{l_1}^2)^r}\prod_{l_2=1}^{p}\frac{C_{2l_2}}{(1+x_{l_1}^2)^r}\prod_{l_3=1}^{q}\frac{C_{3l_3}}{(1+z_{l_1}^2)^r},$$

where $C_{jl_j}, j = 1, 2, 3, l_1 = 1, 2, \ldots, d, l_2 = 1, 2, \ldots, p$ and $l_3 = 1, 2, \ldots, q$ are positive constants such that $C_{l_1}/(1 + y_{l_1}^2)^r$, $C_{l_2}/(1 + x_{l_2}^2)^r$ and $C_{l_3}/(1 + z_{l_3}^2)^r$ are density functions. Define

$$f_n(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) = f_0(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) + c\delta_n^\lambda \prod_{l_1=1}^{d} G\left(\frac{y_{l_1}}{\delta_n}\right) \prod_{l_2=1}^{p} G\left(\frac{x_{l_2}}{\delta_n}\right) \prod_{l_3=1}^{q} G\left(\frac{z_{l_3}}{\delta_n}\right),$$

where the function G is taken to be same as the H function defined in the proof of Theorem 4 from Fan (1991a). Let $F_{\boldsymbol{u}}$ be the joint CDF of $\boldsymbol{U}$, $F_{\boldsymbol{v}}$ be the joint CDF of $\boldsymbol{V}$, and define $g_0 = f_0 * (F_{\boldsymbol{u}} F_{\boldsymbol{v}})$, $g_n = f_n * (F_{\boldsymbol{u}} F_{\boldsymbol{v}})$. Then we have

$$g_0(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) = \prod_{l_1=1}^{d} \frac{C_{1l_1}}{(1 + y_{l_1}^2)^r} \iint \prod_{l_2=1}^{p} \frac{C_{2l_2}}{(1 + (x_{l_2} - u_{l_2})^2)^r} \prod_{l_3=1}^{q} \frac{C_{3l_3}}{(1 + (z_{l_3} - v_{l_3})^2)^r} dF_{\boldsymbol{U}}(\boldsymbol{u}) dF_{\boldsymbol{V}}(\boldsymbol{v})$$

and

$$\begin{aligned} g_n(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) = & \prod_{l_1=1}^{d} \frac{C_{1l_1}}{(1 + y_{l_1}^2)^r} \iint \prod_{l_2=1}^{p} \frac{C_{2l_2}}{(1 + (x_{l_2} - u_{l_2})^2)^r} \prod_{l_3=1}^{q} \frac{C_{3l_3}}{(1 + (z_{l_3} - v_{l_3})^2)^r} dF_{\boldsymbol{U}}(\boldsymbol{u}) dF_{\boldsymbol{V}}(\boldsymbol{v}) \\ & + c\delta_n^\lambda \prod_{l_1=1}^{d} G\left(\frac{y_{l_1}}{\delta_n}\right) \iint \prod_{l_2=1}^{p} G\left(\frac{x_{l_2} - u_{l_2}}{\delta_n}\right) \prod_{l_3=1}^{q} G\left(\frac{z_{l_3} - v_{l_3}}{\delta_n}\right) dF_{\boldsymbol{U}}(\boldsymbol{u}) dF_{\boldsymbol{V}}(\boldsymbol{v}). \end{aligned}$$

Therefore,

$$\begin{aligned} & \iiint [g_n(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) - g_0(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})]^2 g_0^{-1}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{y} d\boldsymbol{x} d\boldsymbol{z} \\ = & \iiint \Bigg[c\delta_n^\lambda \prod_{l_1=1}^{d} G\left(\frac{y_{l_1}}{\delta_n}\right) \cdot \\ & \iint \prod_{l_2=1}^{p} G\left(\frac{x_{l_2} - u_{l_2}}{\delta_n}\right) \prod_{l_3=1}^{q} G\left(\frac{z_{l_3} - v_{l_3}}{\delta_n}\right) dF_{\boldsymbol{U}}(\boldsymbol{u}) dF_{\boldsymbol{V}}(\boldsymbol{v}) \Bigg]^2 g_0^{-1}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{y} d\boldsymbol{x} d\boldsymbol{z} \end{aligned}$$

$$
\begin{aligned}
&= \int \prod_{l_1=1}^{d} G^2\left(\frac{y_{l_1}}{\delta_n}\right)\left[\frac{C_{1l_1}}{(1+y_{l_1}^2)^r}\right]^{-1} d\boldsymbol{y} \cdot \\
&\int \left[c\delta_n^{\lambda}\int \prod_{l_2=1}^{p} G\left(\frac{x_{l_2}-u_{l_2}}{\delta_n}\right) dF_{\boldsymbol{U}}(\boldsymbol{u})\right]^2 \left[\int \prod_{l_2=1}^{p} \frac{C_{2l_2}}{(1+(x_{l_2}-u_{l_2})^2)^r} dF_{\boldsymbol{U}}(\boldsymbol{u})\right]^{-1} d\boldsymbol{x} \cdot \\
&\int \left[\int \prod_{l_3=1}^{q} G\left(\frac{z_{l_3}-v_{l_3}}{\delta_n}\right) dF_{\boldsymbol{V}}(\boldsymbol{v})\right]^2 \left[\int \prod_{l_3=1}^{q} \frac{C_{3l_3}}{(1+(z_{l_3}-v_{l_3})^2)^r} dF_{\boldsymbol{V}}(\boldsymbol{v})\right]^{-1} d\boldsymbol{z}.
\end{aligned}
$$

By changing variables, we have

$$
\begin{aligned}
&\iiint [g_n(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - g_0(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})]^2 g_0^{-1}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) d\boldsymbol{y} d\boldsymbol{x} d\boldsymbol{z} \\
&= c^2 \delta_n^{d+p+q+2\lambda} \int \prod_{l_1=1}^{d} G^2\left(y_{l_1}\right)\left[\frac{C_{1l_1}}{(1+\delta_n^2 y_{l_1}^2)^r}\right]^{-1} d\boldsymbol{y} \cdot \\
&\int \left[\int \prod_{l_2=1}^{p} G\left(x_{l_2}-u_{l_2}\right) dF_{\boldsymbol{U}}(\delta_n \boldsymbol{u})\right]^2 \left[\int \prod_{l_2=1}^{p} \frac{C_{2l_2}}{(1+(\delta_n x_{l_2}-u_{l_2})^2)^r} dF_{\boldsymbol{U}}(\boldsymbol{u})\right]^{-1} d\boldsymbol{x} \cdot \\
&\int \left[\int \prod_{l_3=1}^{q} G\left(z_{l_3}-v_{l_3}\right) dF_{\boldsymbol{V}}(\delta_n \boldsymbol{v})\right]^2 \left[\int \prod_{l_3=1}^{q} \frac{C_{3l_3}}{(1+(\delta_n z_{l_3}-v_{l_3})^2)^r} dF_{\boldsymbol{V}}(\boldsymbol{v})\right]^{-1} d\boldsymbol{z}.
\end{aligned}
$$

To evaluate the first factor in the above expression, we note that, for any constant $M > 0$,

$$
\int G^2(y)(1+\delta_n^2 y^2)^r dy = \int_{|y|>M} G^2(y)(1+\delta_n^2 y^2)^r dy + \int_{|y|\le M} G^2(y)(1+\delta_n^2 y^2)^r dy.
$$

For M large enough, from Condition 3 for G as stated in the proof of Theorem 4 from Fan (1991a), we have $G(y) = O(|y|^{-M_0})$ as $|y| \to \infty$. So,

$$
\int_{|y|>M} G^2(y)(1+\delta_n^2 y^2)^r dy = \int_{|y|>M} c|y|^{-2M_0}(1+\delta_n^2 y^2)^r dy = \int_{|y|>M} c|y|^{-2M_0+2r}(y^{-2}+\delta_n^2)^r dy.
$$

Recall the conditions on M_0 and α, $(M_0+1)\alpha_0 > a > 1+\alpha_0$. This implies that $M_0\alpha_0 > 1$ or $M_0 > 1/\alpha_0$. Choose $r > 0.5$ but very close to 0.5, then we can select an $\alpha_0 \in (0,1)$ so that $1/\alpha_0 > 0.5 + r$, hence $M_0 - r > 1/2$. This leads to $\int_{|y|>M} G^2(y)(1+\delta_n y^2)^r dy = O(1)$ as $\delta_n \to 0$. This, together with the finiteness of $\int_{|y|\le M} G^2(y)(1+\delta_n y^2)^r dy$, implies that $\int G^2(y)(1+\delta_n y^2)^r dy < \infty$ for any δ_n small enough.

Similar to the argument in Fan (1991a) for the ordinary-smooth case, we have

$$
\begin{aligned}
&\int \left[\int \prod_{l_2=1}^{p} G\left(x_{l_2}-u_{l_2}\right) dF_{\boldsymbol{U}}(\delta_n \boldsymbol{u})\right]^2 \left[\int \prod_{l_2=1}^{p} \frac{C_{2l_2}}{(1+(\delta_n x_{l_2}-u_{l_2})^2)^r} dF_{\boldsymbol{U}}(\boldsymbol{u})\right]^{-1} d\boldsymbol{x} \\
&= \int \prod_{l_2=1}^{p} \left[\int G\left(x_{l_2}-u_{l_2}\right) dF_{u_{l_2}}(\delta_n u_{l_2})\right]^2 \left[\int \frac{C_{2l_2}}{(1+(\delta_n x_{l_2}-u_{l_2})^2)^r} dF_{u_{l_2}}(u_{l_2})\right]^{-1} d\boldsymbol{x} \\
&= \prod_{l_2=1}^{p} \int \left[\int G\left(x_{l_2}-u_{l_2}\right) dF_{u_{l_2}}(\delta_n u_{l_2})\right]^2 \left[\int \frac{C_{2l_2}}{(1+(\delta_n x_{l_2}-u_{l_2})^2)^r} dF_{u_{l_2}}(u_{l_2})\right]^{-1} dx_{l_2} \\
&= O(\delta_n^{2[\boldsymbol{\beta}]}),
\end{aligned}
$$

and for the super-smooth case,

$$
\int \left[\int \prod_{l_3=1}^{q} G\left(z_{l_3}-v_{l_3}\right) dF_{\boldsymbol{V}}(\delta_n \boldsymbol{v})\right]^2 \left[\int \prod_{l_3=1}^{q} \frac{C_{3l_3}}{(1+(\delta_n z_{l_3}-v_{l_3})^2)^r} dF_{\boldsymbol{V}}(\boldsymbol{v})\right]^{-1} d\boldsymbol{z}
$$

which is $o\left(\exp\left(-\sum_{l_3=1}^{q} \varepsilon_{l_3}\delta_n^{-\gamma_{l_3}}\right)\right)$ with $\varepsilon_{l_3} = \min\{(1-r)/\gamma_{l_3}, [2(\eta-\eta_0-r)-1]/(2\gamma_{l_3})\} > 0$. Therefore, we have

$$
\iiint \frac{[g_n(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) - g_0(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})]^2}{g_0(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})} d\boldsymbol{y}d\boldsymbol{x}d\boldsymbol{z} = o\left(\delta_n^{2\lambda+2[\boldsymbol{\beta}]+d+p+q} \exp\left(-\sum_{l_3=1}^{q} \varepsilon_{l_3}\delta_n^{-\gamma_{l_3}}\right)\right).
$$

Taking $\delta_n = \varepsilon_1^{1/\gamma_1}(\log n)^{-1/\gamma_1}$ and recall that $\gamma_1 = \max\{\gamma_j : j = 1, 2, \ldots, q\}$, we have

$$\iiint [g_n(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) - g_0(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})]^2 g_0^{-1}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{y} d\boldsymbol{x} d\boldsymbol{z} = o\left(\frac{1}{n}\right).$$

Therefore, for any estimator of $f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0})$ when $f \in \mathscr{C}_{\alpha,m,B}$, we have

$$\sup_{f \in \mathscr{C}_{\alpha,m,B}} E_f[\hat{T}_n - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0})]^2 \geq \frac{c}{4} |f_n^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0}) - f_0^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0})|^2.$$

Note that $f_n^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0}) - f_0^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0})$ can be written as

$$\begin{aligned}
& c\delta_n^{\lambda} \frac{\partial^{[\boldsymbol{k}]}}{\partial \boldsymbol{y}^{\boldsymbol{a}} \partial \boldsymbol{x}^{\boldsymbol{b}} \partial \boldsymbol{z}^{\boldsymbol{c}}} \left[\prod_{l_1=1}^{d} G\left(\frac{y_{l_1}}{\delta_n}\right) \prod_{l_2=1}^{p} G\left(\frac{x_{l_2}}{\delta_n}\right) \prod_{l_3=1}^{q} G\left(\frac{z_{l_3}}{\delta_n}\right) \right] \Bigg|_{\substack{\boldsymbol{y}=\mathbf{0} \\ \boldsymbol{x}=\mathbf{0} \\ \boldsymbol{z}=\mathbf{0}}} \\
= {} & c\delta_n^{\lambda} \left[\prod_{l_1=1}^{d} \frac{\partial^{a_{l_1}}}{\partial y_{l_1}^{a_{l_1}}} G\left(\frac{y_{l_1}}{\delta_n}\right) \prod_{l_2=1}^{p} \frac{\partial^{b_{l_2}}}{\partial x_{l_2}^{b_{l_2}}} G\left(\frac{x_{l_2}}{\delta_n}\right) \prod_{l_3=1}^{q} \frac{\partial^{c_{l_3}}}{\partial z_{l_3}^{c_{l_3}}} G\left(\frac{z_{l_3}}{\delta_n}\right) \right] \Bigg|_{\substack{\boldsymbol{y}=\mathbf{0} \\ \boldsymbol{x}=\mathbf{0} \\ \boldsymbol{z}=\mathbf{0}}} \\
= {} & c\delta_n^{\lambda - [\boldsymbol{k}]} \left[\prod_{l_1=1}^{d} G^{(a_{l_1})}(0) \prod_{l_2=1}^{p} G^{(b_{l_1})}(0) \prod_{l_3=1}^{d} G^{(c_{l_3})}(0) \right],
\end{aligned}$$

which implies that $\sup_{f \in \mathscr{C}_{\alpha,m,B}} E_f[\hat{T}_n - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0})]^2 \geq c\delta_n^{2\lambda - 2[\boldsymbol{k}]}$. Note that $\delta_n = \varepsilon_1^{1/\gamma_1}(\log n)^{-1/\gamma_1}$, also recall that $\lambda = m + \alpha$, so we eventually obtain

$$\sup_{f \in \mathscr{C}_{\alpha,m,B}} E_f[\hat{T}_n - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\mathbf{0}, \mathbf{0}, \mathbf{0})]^2 \geq c(\log n)^{-(2m+2\alpha-2[\boldsymbol{k}])/\gamma_1}$$

for some positive constant c.

If there are no super-smooth measurement error components, then by changing $f_0(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ to $f_0(\boldsymbol{y}, \boldsymbol{x})$, $f_n(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ to $f_n(\boldsymbol{y}, \boldsymbol{x})$, where

$$f_0(\boldsymbol{y}, \boldsymbol{x}) = \prod_{l_1=1}^{d} \frac{C_{1l_1}}{(1 + y_{l_1}^2)^r} \prod_{l_2=1}^{d} \frac{C_{2l_2}}{(1 + x_{l_1}^2)^r}, \quad f_n(\boldsymbol{y}, \boldsymbol{x}) = f_0(\boldsymbol{y}, \boldsymbol{x}) + c\delta_n^{\lambda} \prod_{l_1=1}^{d} G\left(\frac{y_{l_1}}{\delta_n}\right) \prod_{l_2=1}^{p} G\left(\frac{x_{l_2}}{\delta_n}\right),$$

and we will get $\iiint [g_n(\boldsymbol{y},\boldsymbol{x}) - g_0(\boldsymbol{y},\boldsymbol{x})]^2 g_0^{-1}(\boldsymbol{y},\boldsymbol{x}) d\boldsymbol{y} d\boldsymbol{x} = O\left(\delta_n^{2\lambda+2[\boldsymbol{\beta}]+d+p}\right)$. Taking $\delta_n = n^{-1/(2\lambda+2[\boldsymbol{\beta}]+d+p)}$, we have $\iiint [g_n(\boldsymbol{y},\boldsymbol{x}) - g_0(\boldsymbol{y},\boldsymbol{x})]^2 g_0^{-1}(\boldsymbol{y},\boldsymbol{x}) d\boldsymbol{y} d\boldsymbol{x} = O\left(n^{-1}\right)$. Similarly, we can get $\sup_{f\in\mathscr{C}_{\alpha,m,B}} E_f[\hat{T}_n - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\mathbf{0},\mathbf{0})]^2 \geq c\delta_n^{2\lambda-2[\boldsymbol{k}]}$ with $\boldsymbol{k} = (\boldsymbol{a}^T, \boldsymbol{b}^T)$. By plugging $\delta_n = n^{-1/(2\lambda+2[\boldsymbol{\beta}]+d+p)}$ in the above expression, we have

$$\sup_{f\in\mathscr{C}_{\alpha,m,B}} E_f[\hat{T}_n - f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\mathbf{0},\mathbf{0})]^2 \geq cn^{-(2m+2\alpha-2[\boldsymbol{k}])/(2m+2\alpha+2[\boldsymbol{\beta}]+d+p)}.$$

Therefore, we have proven Theorem 2.4.4.

Proof of Theorem 2.4.5. Recall the notation ξ_{nj} defined in (2.4.9) and note that $\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y},\boldsymbol{x}) = n^{-1}\sum_{j=1}^n \xi_{nj}$, by changing variables, we have

$$\begin{aligned}
E\xi_{n1}^2 &= E\left[\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}} \prod_{l_1=1}^d L^{(a_{l_1})}\left(\frac{y_{l_1} - Y_{1l_1}}{h_{l_1}}\right) \prod_{l_2=1}^p H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2} - W_{1l_2}}{\eta_{l_2}}\right)\right]^2 \\
&= \frac{1}{\boldsymbol{h}_\times^{2\boldsymbol{a}+2}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+2}} E\left[\prod_{l_1=1}^d L^{(a_{l_1})}\left(\frac{y_{l_1} - Y_{1l_1}}{h_{l_1}}\right) \prod_{l_2=1}^p \eta_{l_2}^{\beta_{l_2}} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2} - W_{1l_2}}{\eta_{l_2}}\right)\right]^2 \\
&= \frac{1}{\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}} \iint \left[\boldsymbol{\eta}_\times^{\boldsymbol{\beta}} \prod_{l_1=1}^d \prod_{l_2=1}^p L^{(a_{l_1})}(u_{l_1}) H_{nl_2}^{(b_{l_2})}(v_{l_2})\right]^2 f(\boldsymbol{y} - \boldsymbol{u}\boldsymbol{h}, \boldsymbol{x} - \boldsymbol{v}\boldsymbol{\eta}) d\boldsymbol{u} d\boldsymbol{v}.
\end{aligned}$$

Note that

$$\begin{aligned}
&\boldsymbol{\eta}_\times^{\boldsymbol{\beta}} \prod_{l_1=1}^d \prod_{l_2=1}^p L^{(a_{l_1})}(u_{l_1}) H_{nl_2}^{(b_{l_2})}(v_{l_2}) \\
&= \boldsymbol{\eta}_\times^{\boldsymbol{\beta}} \prod_{l_1=1}^d \prod_{l_2=1}^p \left[\frac{1}{2\pi}\int (-is_{l_1})^{a_{l_1}} e^{-is_{l_1}u_{l_1}} \phi_L(s_{l_1}) ds_{l_1}\right] \cdot \\
&\qquad \left[\frac{1}{2\pi}\int (-it_{l_2})^{b_{l_2}} e^{-it_{l_2}v_{l_2}} \phi_H(t_{l_2}) \phi_{u_{l_2}}^{-1}(t_{l_2}/\eta_{l_2}) dt_{l_2}\right],
\end{aligned}$$

and from (O1), for a sufficiently large value M, we have

$$\left|\boldsymbol{\eta}_{\times}^{\boldsymbol{\beta}}\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\frac{\phi_L(s_{l_1})\phi_H(t_{l_2})}{\phi_{u_{l_2}}(t_{l_2}/\eta_{l_2})}\right| = \left|\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\frac{t_{l_2}^{\beta_{l_2}}\phi_L(s_{l_1})\phi_H(t_{l_2})}{(t_{l_2}/\eta_{l_2})^{\beta_{l_2}}\phi_{u_{l_2}}(t_{l_2}/\eta_{l_2})}\right|$$
$$\leq \max\left\{\left|\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\frac{t_{l_2}^{\beta_{l_2}}\phi_L(s_{l_1})\phi_H(t_{l_2})}{(t_{l_2}/\eta_{l_2})^{\beta_{l_2}}\phi_{u_{l_2}}(t_{l_2}/\eta_{l_2})}\right| I(|t_{l_2}/\eta_{l_2}|\geq M),\right.$$
$$\left.\left|\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\frac{t_{l_2}^{\beta_{l_2}}\phi_L(s_{l_1})\phi_H(t_{l_2})}{(t_{l_2}/\eta_{l_2})^{\beta_{l_2}}\phi_{u_{l_2}}(t_{l_2}/\eta_{l_2})}\right| I(|t_{l_2}/\eta_{l_2}|< M)\right\}$$
$$\leq \max\left\{2^p\left|\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\frac{t_{l_2}^{\beta_{l_2}}\phi_L(s_{l_1})\phi_H(t_{l_2})}{c_{l_2}}\right| I(|t_{l_2}|\geq M\eta_{l_2}),\right.$$
$$\left.\left|\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\frac{\phi_L(s_{l_1})\max_{t_{l_2}}|\phi_H(t_{l_2})|}{\min_{|t_{l_2}|\leq M}\phi_{u_{l_2}}(t_{l_2})|}\right| I(|t_{l_2}|< M\eta_{l_2})\right\},$$

which is integrable from (O2). Therefore, by dominated convergence theorem, as $\boldsymbol{\eta}\to 0$, we have

$$\boldsymbol{\eta}_{\times}^{\boldsymbol{\beta}}\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}L^{(a_{l_1})}(u_{l_1})\,H_{nl_2}^{(b_{l_2})}(v_{l_2}) = \frac{1}{(2\pi)^{d+p}}\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\left[\int(-is_{l_1})^{a_{l_1}}e^{-is_{l_1}u_{l_1}}\phi_L(s_{l_1})ds_{l_1}\right]\cdot$$
$$\left[\int(-it_{l_2})^{b_{l_2}}e^{-it_{l_2}v_{l_2}}\eta_{l_2}^{\beta_{l_2}}\phi_H(t_{l_2})\phi_{u_{l_2}}^{-1}(t_{l_2}/\eta_{l_2})dt_{l_2}\right]$$
$$\to\frac{1}{C(2\pi)^{d+p}}\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\left[\int(-is_{l_1})^{a_{l_1}}e^{-is_{l_1}u_{l_1}}\phi_L(s_{l_1})ds_{l_1}\right]\left[\int(-it_{l_2})^{b_{l_2}}e^{-it_{l_2}v_{l_2}}t^{\beta_{l_2}}\phi_H(t_{l_2})dt_{l_2}\right],$$

where $C=c_1\cdots c_p$. Therefore, by the differentiability of f, and the Parseval identity, applying conditions (L1), (O1)–(O3), we have

$$E\xi_{n1}^2 = \frac{1}{\boldsymbol{h}_{\times}^{2\boldsymbol{a}+1}\boldsymbol{\eta}_{\times}^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}}\iint\left[\boldsymbol{\eta}_{\times}^{\boldsymbol{\beta}}\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}L^{(a_{l_1})}(u_{l_1})\,H_{nl_2}^{(b_{l_2})}(v_{l_2})\right]^2 f(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{h},\boldsymbol{x}-\boldsymbol{v}\boldsymbol{\eta})d\boldsymbol{u}d\boldsymbol{v}$$
$$=\frac{f_{\boldsymbol{Y},\boldsymbol{W}}(\boldsymbol{y},\boldsymbol{x})}{(2\pi)^{d+p}C^2\boldsymbol{h}_{\times}^{2\boldsymbol{a}+1}\boldsymbol{\eta}_{\times}^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}}\iint\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}|s_{l_1}|^{2a_{l_1}}|t_{l_2}|^{2b_{l_2}+2\beta_{l_2}}|\phi_L(s_{l_1})|^2|\phi_H(t_{l_2})|^2d\boldsymbol{s}d\boldsymbol{t}\cdot[1+o(1)].$$

Similarly, we can show that $E|\xi_{n1}|^{2+\delta} = O\left(\boldsymbol{h}_\times^{-(2+\delta)(\boldsymbol{a}+1)+1}\boldsymbol{\eta}_\times^{-(2+\delta)(\boldsymbol{b}+1)-(2+\delta)\boldsymbol{\beta}+1}\right)$ for any positive δ. This, together with the result $E\xi_{n1} \to f_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x})$, we have

$$\frac{nE|\xi_{n1} - E\xi_{n1}|^{2+\delta}}{[n\mathrm{Var}(\xi_{n1})]^{(2+\delta)/2}} = O\left(\frac{n(\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1})^{(2+\delta)/2}}{n^{(2+\delta)/2}\boldsymbol{h}_\times^{(2+\delta)(\boldsymbol{a}+1)-1}\boldsymbol{\eta}_\times^{(2+\delta)(\boldsymbol{b}+1)+(2+\delta)\boldsymbol{\beta}-1}}\right) = O\left(\frac{1}{(\sqrt{n\boldsymbol{h}_\times\boldsymbol{\eta}_\times})^\delta}\right)$$

which is the order of $o(1)$ as $n\boldsymbol{h}_\times\boldsymbol{\eta}_\times \to \infty$ as assumed. Hence, by Lyapunov central limit theorem, we proved Theorem 2.4.5.

Proof of Theorem 2.4.6:

As we discussed before, it is difficult to find the variance of $\hat{f}_{\boldsymbol{D}}^{(\boldsymbol{k})}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$. To apply the Lyapunov central limit theorem, we have to find a lower bound of $E\xi_{n1}^2$, where ξ_{n1} can be written as

$$\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\left[\prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1} - Y_{1l_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2} - W_{1l_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3} - S_{1l_3}}{\rho_{l_3}}\right)\right],$$

and $L^{(a_{l_1})}(y) = (2\pi)^{-1}\int(-it)^{a_{l_1}}\exp(-ity)\phi_L(t)dt$,

$$\begin{aligned}
H_{nl_2}^{(b_{l_2})}(x) &= \frac{1}{2\pi}\int(-it)^{b_{l_2}}\exp(-itx)\phi_H(t)/\phi_{U_{l_2}}(t/\eta_{l_2})dt, \\
K_{nl_3}^{(c_{l_3})}(z) &= \frac{1}{2\pi}\int(-it)^{c_{l_3}}\exp(-itz)\phi_K(t)/\phi_{V_{l_3}}(t/\rho_{l_3})dt.
\end{aligned}$$

By changing of variables, one can see that $E\xi_{n1}^2$ equals

$$\begin{aligned}
&E\left[\frac{1}{\boldsymbol{h}_\times^{\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{\boldsymbol{c}+1}}\prod_{l_1=1}^{d} L^{(a_{l_1})}\left(\frac{y_{l_1} - Y_{jl_1}}{h_{l_1}}\right)\prod_{l_2=1}^{p} H_{nl_2}^{(b_{l_2})}\left(\frac{x_{l_2} - W_{jl_2}}{\eta_{l_2}}\right)\prod_{l_3=1}^{q} K_{nl_3}^{(c_{l_3})}\left(\frac{z_{l_3} - S_{1l_3}}{\rho_{l_3}}\right)\right]^2 \\
&= \frac{1}{\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+1}\boldsymbol{\rho}_\times^{2\boldsymbol{c}+1}}\iiint\left[\prod_{l_1=1}^{d}\prod_{l_2=1}^{p}\prod_{l_3=1}^{q} L^{(a_{l_1})}(u_{l_1})\, H_{nl_2}^{(b_{l_2})}(w_{l_2})\, K_{nl_3}^{(c_{l_3})}(s_{l_3})\right]^2 \cdot \\
&\qquad f(\boldsymbol{y} - \boldsymbol{u}\boldsymbol{h}, \boldsymbol{x} - \boldsymbol{w}\boldsymbol{\eta}, \boldsymbol{z} - \boldsymbol{s}\boldsymbol{\rho})d\boldsymbol{u}d\boldsymbol{v}d\boldsymbol{s}.
\end{aligned}$$

For the ordinary smooth measurement error, we have

$$\eta_n^{\beta_{l_2}} H_{nl_2}^{(b_{l_2})}(w_{l_2}) \to \frac{1}{2\pi c}\int (-it)^{b_{l_3}} \exp(-itw_{l_2}) t^{\beta_{l_2}} \phi_H(t)dt, \quad \left|\eta_n^{\beta_{l_2}} H_{nl_2}^{(b_{l_2})}(w_{l_2})\right| \le \min(d_2, d_1/w_{l_2})$$

for some positive constants d_1 and d_2; for the super-smooth part, from Fan (1991b),we have

$$\left|K_{nl_3}^{(c_{l_3})}(s_{l_3})\right| \ge d_3 \left|\frac{d^{(c_{l_3})}\cos(s_{l_3})}{ds_{l_3}^{c_{l_3}}}\right| \rho_{l_3}^{\mu_{l_3 0}} b_{nl_3}^{m+4} \exp\left((1-b_{nl_3})^{\gamma_{l_3}}/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}})\right),$$

where $b_{nl_3} = \rho_{l_3}^{\mu_{l_3 0}/2(m+5)}$ for $s_{l_3} \in [0, \pi/2]$.

Therefore, when n is sufficiently large, by Parseval identity,

$$\begin{aligned}
E\xi_{n1}^2 \ge & \frac{cf(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})}{\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}\boldsymbol{\rho}_\times^{2\boldsymbol{c}+1}} \cdot \\
& \iint \prod_{l_1}^{d} [L^{(a_{l_1})}(u_{l_1})]^2 \prod_{l_2=1}^{p} \left[\frac{1}{2\pi}\int \exp(-itw_{l_2})(-it)^{b_{l_2}} t^{\beta_{l_2}}\phi_H(t)dt\right]^2 d\boldsymbol{u}d\boldsymbol{w} \cdot \\
& \int_0^{\pi/2} \prod_{l_3=1}^{q} \left[\rho_{l_3}^{\mu_{l_3 0}} b_n^{m+4} \exp\left((1-b_n)^{\gamma_{l_3}}/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}})\right)\right]^2 d\boldsymbol{v} \\
= & \frac{cf(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})}{\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}\boldsymbol{\rho}_\times^{2\boldsymbol{c}+1}} \int \prod_{l_1}^{d} t_{l_1}^{2a_{l_1}} \phi_L^2(t_{l_1}) d\boldsymbol{t} \cdot \int \prod_{l_2=1}^{p} t_{l_2}^{2b_{l_2}+2\beta_{l_2}} \phi_H^2(t_{l_2}) d\boldsymbol{t} \cdot \\
& \prod_{l_3=1}^{q} \left[\rho_{l_3}^{\mu_{l_3 0}} b_{nl_3}^{m+4} \exp\left((1-b_{nl_3})^{\gamma_{l_3}}/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}})\right)\right]^2 \\
\ge & \frac{cf(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\boldsymbol{\rho}_\times^{\boldsymbol{\mu}_0[1+(m+4)/(2(m+5))]-2\boldsymbol{c}-1}}{\boldsymbol{h}_\times^{2\boldsymbol{a}+1}\boldsymbol{\eta}_\times^{2\boldsymbol{b}+2\boldsymbol{\beta}+1}} \prod_{l_3=1}^{q} \exp\left(\frac{2-4\gamma_{l_3} b_{nl_3}}{\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}}}\right),
\end{aligned}$$

where $\boldsymbol{\mu}_0 = (\mu_{10}, \ldots, \mu_{q0})$. On the other hand, by the definition of $K_{nl_3}^{(c_{l_3})}(s)$, we have

$$\begin{aligned}
|K_{nl_3}^{(c_{l_3})}(s)| \le & \frac{1}{2\pi}\left|\int_{-1}^{1} (-it)^{c_{l_3}} \exp(-its)\phi_K(t)/\phi_{v_{l_3}}(t/\rho_{l_3})dt\right| \\
\le & \frac{1}{2\pi}\int_{|t|>M} |t|^{\xi_{l_3 0}} \exp(|t|^{\gamma_{l_3}}/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}}))dt + O(1) = O\left(\exp(1/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}}))\right).
\end{aligned}$$

This, together with the result from the ordinary smooth, we have, for $\delta > 0$,

$$E|\xi_{n1}|^{2+\delta} = O\left(\boldsymbol{h}_{\times}^{1-(2+\delta)(\boldsymbol{a}+1)}\boldsymbol{\eta}_{\times}^{1-(2+\delta)(\boldsymbol{b}+1)-(2+\delta)\boldsymbol{\beta}}\boldsymbol{\rho}_{\times}^{-(2+\delta)(\boldsymbol{c}+1)}\prod_{l_3=1}^{q}\exp((2+\delta)/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}}))\right).$$

Therefore, after a tedious algebra, we have

$$\frac{E|\xi_{n1}|^{2+\delta}}{n^{\delta/2}(E|\xi_{n1}|^2)^{(2+\delta)/2}} = O\left(\frac{1}{(\sqrt{n\boldsymbol{\eta}_{\times}\boldsymbol{h}_{\times}})^{\delta}\boldsymbol{\rho}^{-c(2+\delta)}}\prod_{l_3=1}^{q}\exp(2(2+\delta)\gamma_{l_3}b_{nl_3}/(\xi_{l_3}\rho_{l_3}^{\gamma_{l_3}}))\right).$$

Note that $b_{nl_3} = \rho_{l_3}^{\gamma_{l_3}/(2(m+5))}$, and $-1 < -1 + 1/(2(m+5)) < 0$, so based on the assumption (S3), we can show that the Lyapunov condition holds. This concludes the proof of Theorem 2.4.6.

Chapter 3

Estimating Density Function of Data Mixing with Berkson Error

3.1 Introduction

In the measurement error literature, the classical structure has received more attention from researchers than the Berkson structure. To be specific, suppose the random variable of interest is X, but for some reason, we cannot directly observe its value. Instead, what we can observe is a random variable Z which relates to X in two different ways as generally assumed in literature. One is the classic error which assumes that $Z = X + U$, where X and U are independent, another is the Berkson error to assume that $X = Z + U$, where Z and U are independent. In both structures, U is centered at 0 and called the measurement error. However, the different dependence structures in classic and Berkson measurement error result in significantly different statistical inference procedures.

Since the first research on Berkson measurement errors conducted by Berkson (1950) in the regression context, the Berkson measurement error structure has been widely practiced in designed experiments in which there are some target values that an experimenter wants to apply to experimental units, but the actual delivered values may differ from the target values. For example, a nominal amount of nitrogen (Z) is applied to a crop, but the actual amount of nitrogen (X) absorbed by the crop could be different from the nominal one and is difficult to measure. In this scenario, the relationship between the actual amount of nitrogen and the nominal one is better explained by a Berkson measurement error structure. Many other examples of real data applications with Berkson measurement errors can be found in many areas including agriculture experiments, bioassay studies, and epidemiology, among others, as described in the monograph Carroll et al. (2006), journal papers Carroll et al. (2007b), Delaigle (2007) and Long et al. (2016), and the references therein.

To our best knowledge, it seems that the statistical research related to the Berkson error has not gained much attention from statisticians and practitioners. The main reason might be that the Berkson measurement error in parametric models may not be as challenging as in classical measurement error setups. A simple example to this point is that the naive least square estimator in the linear regression with the Berkson error is unbiased. Even in nonlinear regression models, the bias caused by Berkson measurement error can be easily attenuated by regressing the response variable on the noisy variables if the distribution of the measurement error is known, a common assumption in measurement error modeling literature. See Wang (2003, 2004, 2007) for more details. However, the statistical inference related to the Berkson measurement errors appears much more difficult. Recent years have evidenced increased interest in developing nonparametric inference procedures, such as the nonparametric density estimators. Delaigle (2007) proposed a simple density estimation procedure while studying a more comprehensive setting including both Berkson and classical errors. Later, Long et al. (2016) discusses the bandwidth selection issues in kernel based estimation procedures.

In real applications, we might be interested in modeling the relationship among many variables. Due to the differences in measuring instruments, some variables can be measured directly without measurement error, and some variables may carry measurement errors, such as the Berkson error. Further data analysis, therefore, should take care of these errors in order to have a reliable inference procedures. In this paper, we would like to propose estimation procedures to estimate the joint density function of two random variables, with one variable being error-free, and another variable contaminated with Berkson error. Two estimation procedures, based on whether applying smoothing to the variable with Berkson error, will be considered. The estimation of joint density function has its own theoretical and practical interest, but the resulting density estimator can be also used as building block to construct the estimator for the regression functions when predictors have both types of variables in the model.

The paper is organized as follows. Two nonparametric estimators of the joint density function are proposed in Section 2. Technical assumptions, mean squared errors (MSE), and large sample theoretical properties are discussed in Section 3. A discussion on the optimal convergence rates and the bandwidth choice of the proposed estimators is conducted in Section 4. In addition, the performance of the developed data-driven bandwidth selection method is evaluated by a simulation study. All the technical proofs of the theoretical results are deferred to Section 5.

3.2 Estimation Procedures

Consider a random pair (X, Y), where X can not be observed directly, one can observe Z such that $X = Z + U$. That is, X is contaminated with a Berkson type of error. For identifiability, the density function of the measurement error U is assumed to be known. Y can be directly observed. In this paper, we will propose two estimators for the joint density function of (X, Y) based on a sample (Z_j, Y_j), $j = 1, 2, \ldots, n$, from (Z, Y).

Note that the independence between Z and U in $X = Z+U$ implies that we can estimate the density function of X by the simple average $\hat{f}_X(x) = n^{-1} \sum_{j=1}^{n} f_u(x - Z_j)$, which is first proposed and discussed in Delaigle (2007) by exploiting the Berkson structure. Combining this component with the classical kernel smoothing idea for the observed data on Y, we define an estimator for the joint density of (X, Y) as

$$\tilde{f}_{X,Y}(x, y) = \frac{1}{nh} \sum_{j=1}^{n} f_u(x - Z_j) K\left(\frac{y - Y_j}{h}\right). \qquad (3.2.1)$$

The estimator proposed in (3.2.1) is very simple to calculate and does not need any specification of smoothing parameters for the Berkson component, which is very different from the standard kernel density estimator in the error-free case or the deconvolution kernel estimator in the classical measurement error case. However, as noted in Long et al. (2016), a disadvantage of $\hat{f}_X(x)$ is that in cases where U is concentrated around 0, the estimator will have large spikes at the sample points Z_i's and thus have high variability.

To overcome the drawback in the construction (3.2.1), we propose an estimator of the joint density function of (X, Y) by smoothing the Berkson measurement error component. Note that the empirical characteristic function of Z is defined as $\phi_n(t) = n^{-1}\sum_{j=1}^n e^{itZ_j}$, where i is the imaginary unit such that $i^2 = -1$. Thus, from $X = Z + U$ and the independence of Z and U, we have $\phi_X(t) = \phi_Z(t)\phi_U(t)$. Choosing K as a proper kernel density function and h a sequence positive numbers tending to 0 as $n \to \infty$, we can estimate the density function of Z by the classical kernel estimator having the form of $\hat{f}_Z(z) = (nh)^{-1}\sum_{j=1}^n K((z - Z_j)/h)$. Therefore, an estimator of the characteristic function of X can be defined as

$$\hat{\phi}_X(t) = \hat{\phi}_Z(t)\phi_U(t) = \phi_K(th)\phi_U(t)\phi_n(t).$$

Assuming that $\hat{\phi}_X(t) \in L_1$, we may estimate f_X using $\hat{f}_X(x) = \frac{1}{2\pi}\int e^{-itx}\hat{\phi}_X(t)dt$, or

$$\begin{aligned}\hat{f}_X(x) &= \frac{1}{2\pi}\int e^{-itx}\phi_K(th)\phi_U(t)\phi_n(t)dt = \frac{1}{2\pi}\int e^{-itx}\frac{1}{n}\sum_{j=1}^n e^{itZ_j}\phi_K(th)\phi_U(t)dt \qquad (3.2.2)\\ &= \frac{1}{n}\sum_{j=1}^n \frac{1}{2\pi}\int e^{-it(x-Z_j)}\phi_K(th)\phi_U(t)dt = \frac{1}{nh}\sum_{j=1}^n \int K\left(\frac{x-u-Z_j}{h}\right) f_U(u)du.\end{aligned}$$

The estimator $\hat{f}_X(x)$ was proposed in Long et al. (2016). Therefore, an estimator of joint density function of (X, Y) can be constructed as

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh_1h_2}\sum_{j=1}^n L\left(\frac{y-Y_j}{h_2}\right)\int K\left(\frac{x-u-Z_j}{h_1}\right) f_U(u)du, \qquad (3.2.3)$$

where L and K are two kernel functions, and h_1, h_2 are two bandwidths. A connection between the estimator (3.2.2) and the simple average estimator proposed in Delaigle (2007) is that if we let $h \to 0$, then the continuity of f_U implies that the latter is the limit of the former.

3.3 Large Sample Results

In this section, the finite sample and large sample properties of the estimators proposed in (3.2.1) and (3.2.3) will be investigated. We begin with the discussion on the mean squared errors (MSE) for (3.2.1). Denote $R(K) = \int K^2(x)dx$, $f''_{Z,Y}(z,y)$ the second derivative of $f_{Z,Y}(z,y)$ with respect to y. We also need the following technical conditions to validate the theoretical results.

(C1). (Z, Y) and U are independent.

(C2). The joint density function $f_{Z,Y}(z,y)$ is twice continuously differentiable with respect to y.

(C3). The kernel density function K is bounded and satisfies $\int K(x)dx = 1$, $\int xK(x)dx = 0$, and $\int x^2K(x)dx = 1$.

(C4). The bandwidth h is so chosen that $h \to 0$ and $nh \to \infty$.

(C5). For each (x, y) in the domain of (X, Y), there is a positive number $\gamma > 0$,

$$\int f_U^{2+\delta}(x-z)[f_{Z,Y}(z,y) + |f'_{Z,Y}(z,y)| + |f''_{Z,Y}(z,y)|]dz < \infty.$$

(C6). The bandwidth h is so chosen that $h \to 0$ and $nh/\log n \to \infty$.

(C7). The joint density function $f_{X,Y}(x,y)$ is uniformly-continuous.

(C8). The kernel function K is bounded and satisfies $\int K(x)dx = 1$, $\int xK(x)dx = 0$, and $\int x^2K(x)dx = 1$. It also has an integrable characteristic function.

(C9). The bandwidth h is so chosen that $h \to 0$, $nh^2 \to \infty$.

(D1). The kernel function K and L is bounded and satisfies $\int K(x)dx = \int L(x)dx = 1$, $\int xK(x)dx = \int xL(x)dx = 0$, and $\int x^2K(x)dx = \int x^2L(x)dx = 1$. They also have integrable characteristic functions.

(D2). The bandwidths h_1, h_2 is so chosen that $h_1, h_2 \to 0$ and $nh_2 \to \infty$.

(D3). The joint density function $f_{X,Y}(x, y)$ is twice continuously differentiable with respect to x and y.

(D4). The bandwidths h_1, h_2 are so chosen that $h_1, h_2 \to 0$ and $nh_1h_2/\log n \to \infty$.

(D5). The bandwidth h_1, h_2 are so chosen that $h_1, h_2 \to 0$, $nh_2^2 \to \infty$.

Conditions (C1)-(C4) are needed to evaluate the MSE of $\tilde{f}_{X,Y}(x, y)$. Together with the condition (C5), the asymptotic normality of $\tilde{f}_{X,Y}(x, y)$ can be derived. (C6) is required to show the almost sure convergence, and with the additional assumptions (C8), (C9), a uniform convergence in probability result can be established for $\tilde{f}_{X,Y}(x, y)$. A parallel set of conditions (D1)-(D5) are imposed to validate all the large sample results for the estimator $\hat{f}_{X,Y}(x, y)$. These conditions are mild and also widely adopted in the kernel smoothing literature.

We begin with the introduction of the MSE expression for $\tilde{f}_{X,Y}(x, y)$.

Theorem 3.3.1. *Assume the conditions (C1)-(C4) hold. For $\tilde{f}_{X,Y}(x, y)$ defined in (3.2.1),*

$$\begin{aligned} MSE(\tilde{f}_{X,Y}(x,y)) = & \frac{R(K)}{nh} \int f_U^2(x-z) f_{Z,Y}(z,y) dz + \\ & + \frac{h^4}{4} \left[\int f_U(x-z) f''_{Z,Y}(z,y) dz \right]^2 + o\left(\frac{1}{nh}\right) + o(h^4). \end{aligned} \tag{3.3.1}$$

An optimal MSE convergence rate can be obtained by choosing an optimal bandwidth h to minimize the asymptotic MSE (AMSE) expression in Theorem 3.3.1. The AMSE is defined as the first two terms in (3.3.1). In fact, by setting the derivatives of $AMSE$ equal to 0, we obtain

$$h_{\text{opt}} = \frac{1}{n^{1/5}} \left(\frac{R(K) \int f_U^2(x-z) f_{Z,Y}(z,y) dz}{\left[\int f_U(x-z) f''_{Z,Y}(z,y) dz\right]^2} \right)^{1/5}.$$

Accordingly, the optimal convergence rate of MSE, by plugging h_{opt} into the expression of MSE in (3.3.1) is

$$\mathrm{MSE}(\tilde{f}_{X,Y}(x,y)) = \frac{5}{4n^{4/5}}\left(R(K)\int f_U^2(x-z)f_{Z,Y}(z,y)dz\right)^{4/5}\left(\int f_U(x-z)f''_{Z,Y}(z,y)dz\right)^{2/5}.$$

It is noted that the optimal convergence rate is the same as the one for the classical univariate kernel density estimator. The theoretical optimal bandwidth depends on unknown quantities, and therefore cannot be implemented for real data analysis. For implementation, some data driven bandwidth selectors should be considered, such as the cross validation procedures discussed in Wand and Jones (1995).

The next theorem states the asymptotic normality of $\tilde{f}_{X,Y}(x,y)$.

Theorem 3.3.2. *In addition to the conditions in Theorem 3.3.1, we further assume that (C5) holds. Then*

$$\sqrt{nh\delta(x,y)}\cdot\left[\tilde{f}_{X,Y}(x,y) - f_{X,Y}(x,y) - \frac{h^2}{2}\int f_U(x-z)f''_{Z,Y}(z,y)dz + o(h^2)\right] \Longrightarrow N(0,1),$$

where $\delta(x,y) = (R(K)\int f_U^2(x-z)f_{Z,Y}(z,y)dz)^{-1}$. *If we further assume that* $nh^5 = o(1)$, *then*

$$\sqrt{nh\delta(x,y)}\left[\tilde{f}_{X,Y}(x,y) - f_{X,Y}(x,y)\right] \Longrightarrow N(0,1).$$

The following theorem indicates that, under some stronger conditions, $\tilde{f}_{X,Y}(x,y)$ converges to $f_{X,Y}(x,y)$ almost surely.

Theorem 3.3.3. *Suppose the conditions (C1)-(C3) and (C6) hold. Then* $\tilde{f}_{X,Y}(x,y)$ *converges to* $f_{X,Y}(x,y)$ *almost surely as* $n \to \infty$.

By imposing some conditions on the smoothness of the joint density function $f_{X,Y}(x,y)$, we can show that $\tilde{f}_{X,Y}(x,y)$ is uniformly convergent in probability.

Theorem 3.3.4. *Suppose (C1), (C2), (C7)-(C9) hold. Then*

$$\sup_{x,y}|\tilde{f}_{X,Y}(x,y) - f_{X,Y}(x,y)| = o_p(1).$$

Now let's move our attention to the estimator $\hat{f}_{X,Y}(x,y)$ defined 3.2.3. We would like derive a series of finite and large sample results similar to those established for $\tilde{f}_{X,Y}(x,y)$. We begin with a theorem on the MSE of $\hat{f}_{X,Y}(x,y)$.

Theorem 3.3.5. *Suppose (D1)-(D3) hold. Then*

$$\begin{aligned} MSE(\hat{f}_{X,Y}(x,y)) &= \frac{h_1^4}{4}\left[\frac{\partial f^2_{X,Y}(x,y)}{\partial x^2}\right]^2 + \frac{h_2^4}{4}\left[\frac{\partial f^2_{X,Y}(x,y)}{\partial y^2}\right]^2 \\ &+ \frac{R(L)}{nh_2}\iint f_U^2(x-z)f_{Z,Y}(z,y)dz + o\left(\frac{1}{nh_2}\right) + o((h_1^2+h_2^2)^2). \end{aligned} \tag{3.3.2}$$

It is interesting to note that the optimal bandwidths h_1, h_2, which minimize the AMSE, or the first three terms in the expression of $\mathrm{MSE}(\hat{f}_{X,Y}(x,y))$, are $h_{1,\mathrm{opt}} = 0$,

$$h_{2,\mathrm{opt}} = \frac{1}{n^{1/5}}\left(\frac{R(L)\int f_U^2(x-z)f_{Z,Y}(z,y)dz}{\left[\partial^2 f_{X,Y}(x,y)/\partial y^2\right]^2}\right)^{1/5},$$

and the optimal convergence rate of MSE, by plugging $h_{1,\mathrm{opt}} = 0$, $h_{2,\mathrm{opt}}$ into the expression of MSE in (3.3.2) is

$$\mathrm{MSE}(\hat{f}_{X,Y}(x,y)) = \frac{5}{4n^{4/5}}\left(R(L)\int f_U^2(x-z)f_{Z,Y}(z,y)dz\right)^{4/5}\left(\partial^2 f_{X,Y}(x,y)/\partial y^2\right)^{2/5}.$$

Note that $\partial^2 f_{X,Y}(x,y)/\partial y^2 = \int f_U(x-z)f''_{Z,Y}(z,y)dz$, it is easy to see the optimal convergence rates of $\tilde{f}_{X,Y}(x,y)$ and $\hat{f}_{X,Y}(x,y)$ are exactly same.

The following theorem claims that $\hat{f}_{X,Y}(x,y)$ is asymptotically normal.

Theorem 3.3.6. *Suppose (C5) and (D1)-(D3) hold. Then*

$$\sqrt{nh_2\delta(x,y)}\left[\hat{f}_{X,Y}(x,y) - f_{X,Y}(x,y) - \frac{h_1^2}{2}\frac{\partial f^2_{X,Y}(x,y)}{\partial x^2} + \frac{h_2^2}{2}\frac{\partial f^2_{X,Y}(x,y)}{\partial y^2} + o(h_1^2) + o(h_2^2)\right]$$

is asymptotically standard normal, where $\delta(x,y) = (R(L)\int f_U^2(x-z)f_{Z,Y}(z,y)dz)^{-1}$. *If we* *further assume that $nh_1^4 h_2 \to 0$ and $nh_2^5 \to 0$, then we have*

$$\sqrt{nh_2\delta(x,y)}\left[\hat{f}_{X,Y}(x,y) - f_{X,Y}(x,y)\right] \Longrightarrow N(0,1).$$

The following theorem claims that $\hat{f}_{X,Y}(x,y)$ also converges almost surely.

Theorem 3.3.7. *Suppose (D1)-(D4) hold. Then $\hat{f}_{X,Y}(x,y)$ converges to $f_{X,Y}(x,y)$ almost surely.*

Similar to the estimator $\tilde{f}_{X,Y}(x,y)$, after imposing some conditions on the characteristic functions of the kernel functions, we can show that $\hat{f}_{X,Y}(x,y)$ is uniformly convergent in probability.

Theorem 3.3.8. *Suppose (C4), (D1), (D3), (D5) hold. Then*

$$\sup_{x,y} |\hat{f}_{X,Y}(x,y) - f_{X,Y}(x,y)| = o_p(1).$$

3.4 Discussion

We conjecture that the two estimators proposed in this chapter attains the fast convergence rate. A non-rigorous justification of this claim is based on the following observations: Chang and Wu (2015) once pointed out that "Usually the convergence rates of multivariate density estimations are slower than those of univariate densities", and we also know that under some very mild conditions, the classic Rosenblatt-Parzen kernel density estimator achieves fast convergence rate among all nonparametric estimators of a density function. Based on the asymptotic normality established in Theorem 3.3.3 and 3.3.6, we know the convergence rates of $\tilde{f}_{X,Y}(x,y)$ and $\hat{f}_{X,Y}(x,y)$ are the same as the convergence rate of a univariate kernel density estimate.

It is well known that the bandwidth has a significant impact on the performance of a kernel density estimator. Because of the presence of the unknown quantities, the optimal bandwidths of $\tilde{f}_{X,Y}(x, y)$ and $\hat{f}_{X,Y}(x, y)$ derived in Section 3.3 based on Theorem 3.3.1 and 3.3.5 cannot be used directly for estimating the density function of (X, Y). Data-driven bandwidth selectors are often preferred in implementing various smoothing procedures. In the following, we will derive a least squares cross validation (LSCV) bandwidth selector for $\tilde{f}_{X,Y}(x, y)$. Note that the mean integrated squared error (MISE) of $\tilde{f}_{X,Y}$, defined as $E \iint [\tilde{f}_{X,Y}(x, y) - f_{X,Y}(x, y)]^2 dxdy$, equals

$$E \iint \tilde{f}^2_{X,Y}(x, y)dxdy - 2E \iint \tilde{f}_{X,Y}(x, y)f_{X,Y}(x, y)dxdy + \iint f^2_{X,Y}(x, y)dxdy,$$

and $\iint f^2_{X,Y}(x, y)dxdy$ is free of h, so we only have to consider the first two terms in the above expression. The first expectation can be estimated by

$$\begin{aligned}
&\iint \tilde{f}^2_{X,Y}(x, y)dxdy = \iint \left[\frac{1}{nh}\sum_{j=1}^{n} f_u(x - Z_j)K\left(\frac{y - Y_j}{h}\right)\right]^2 dxdy \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \iint f_u(x - Z_i)K\left(\frac{y - Y_i}{h}\right) f_u(x - Z_j)K\left(\frac{y - Y_j}{h}\right) dxdy \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left[\int f_u(x - Z_i)f_u(x - Z_j)dx\right]\left[\int K\left(\frac{y - Y_i}{h}\right) K\left(\frac{y - Y_j}{h}\right) dy\right].
\end{aligned}$$

To find an estimator for the second expectation, we note that

$$\begin{aligned}
&\iint \tilde{f}_{X,Y}(x, y)f_{X,Y}(x, y)dxdy = \iint \tilde{f}_{X,Y}(x, y)\left[\int f_{Z,Y}(z, y)f_u(x - z)dz\right] dxdy \\
&= \iint \left[\int \tilde{f}_{X,Y}(x, y)f_u(x - z)dx\right] f_{Z,Y}(z, y)dzdy \\
&= \iint \left[\int \frac{1}{nh}\sum_{j=1}^{n} f_u(x - Z_j)K\left(\frac{y - Y_j}{h}\right) f_u(x - z)dx\right] f_{Z,Y}(z, y)dzdy
\end{aligned}$$

$$= \iint \frac{1}{nh} \sum_{j=1}^{n} \left[\int f_u(x - Z_j) f_u(x - z) dx \right] K\left(\frac{y - Y_j}{h} \right) f_{Z,Y}(z, y) dz dy$$
$$= \iint \hat{g}(z, y) f_{Z,Y}(z, y) dz dy,$$

where
$$\hat{g}(z, y) = \frac{1}{nh} \sum_{j=1}^{n} \left[\int f_u(x - Z_j) f_u(x - z) dx \right] K\left(\frac{y - Y_j}{h} \right).$$

Then an unbiased estimator of the second expectation is given by $n^{-1} \sum_{i=1}^{n} \hat{g}_{-i}(Z_i, Y_i)$, where, for each $i = 1, 2, \ldots, n$,

$$\hat{g}_{-i}(Z_i, Y_i) = \frac{1}{n-1} \sum_{j \neq i} K\left(\frac{Y_i - Y_j}{h} \right) \int f_u(x - Z_j) f_u(x - Z_i) dx.$$

Therefore, the LSCV function to be minimized has the form of

$$\begin{aligned} \mathrm{LSCV}(h) = & \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\int f_u(x - Z_i) f_u(x - Z_j) dx \right] \left[\int K\left(\frac{y - Y_i}{h} \right) K\left(\frac{y - Y_j}{h} \right) dy \right] \\ & - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{Y_i - Y_j}{h} \right) \int f_u(x - Z_j) f_u(x - Z_i) dx. \end{aligned}$$

In particular, if the density function f_u of U is normal, and the kernel function K is chosen to be standard normal, then the LSCV(h) function can be simplified to

$$\begin{aligned} \mathrm{LSCV}(h) = & \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi(Z_i - Z_j, 0, \sqrt{2}\sigma_u) \phi(Y_i - Y_j, 0, \sqrt{2}h) \\ & - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \phi(Y_i - Y_j, 0, \sqrt{h}) \phi(Z_i - Z_j, 0, \sqrt{2}\sigma_u). \end{aligned} \tag{3.4.1}$$

where $\phi(v, 0, \sigma)$ is the normal density function with mean 0 and standard deviation σ evaluated at v.

Similarly, a LSCV bandwidth selector can also be constructed for $\hat{f}_{X,Y}(x, y)$. If U has a normal distribution $N(0, \sigma_u^2)$ and both K and L are chosen to be standard normal, then

$$\iint \hat{f}_{X,Y}^2(x,y)dxdy = \iint \left[\frac{1}{nh_1h_2} \sum_{j=1}^{n} L\left(\frac{y-Y_j}{h_2}\right) \int K\left(\frac{x-u-Z_j}{h_1}\right) f_u(u)du \right]^2 dxdy$$

$$= \frac{1}{n^2h_1^2h_2^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \int L\left(\frac{y-Y_i}{h_2}\right) L\left(\frac{y-Y_j}{h_2}\right) dy \cdot$$

$$\int \left[\int K\left(\frac{x-u-Z_i}{h_1}\right) f_u(u)du\right]\left[\int K\left(\frac{x-u-Z_j}{h_1}\right) f_u(u)du\right] dx$$

$$= \frac{1}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \phi(Y_i - Y_j, 0, \sqrt{2}h_2) \int \phi\left(x - Z_i, 0, \sqrt{h_1^2 + \sigma_u^2}\right) \phi\left(x - Z_j, 0, \sqrt{h_1^2 + \sigma_u^2}\right) dx$$

$$= \frac{1}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \phi(Y_i - Y_j, 0, \sqrt{2}h_2) \phi\left(Z_i - Z_j, 0, \sqrt{2(h_1^2 + \sigma_u^2)}\right).$$

We also have

$$\iint \hat{f}_{X,Y}(x,y) f_{X,Y}(x,y)dxdy = \iint \hat{f}_{X,Y}(x,y) \left[\int f_{Z,Y}(z,y) f_u(x-z)dz\right] dxdy$$

$$= \iint \left[\int \hat{f}_{X,Y}(x,y) f_u(x-z)dx\right] f_{Z,Y}(z,y)dzdy$$

$$= \iint \left[\int \frac{1}{nh_1h_2} \sum_{j=1}^{n} L\left(\frac{y-Y_j}{h_2}\right) \left(\int K\left(\frac{x-u-Z_j}{h_1}\right) f_u(u)du\right) f_u(x-z)dx\right] f_{Z,Y}(z,y)dzdy$$

$$= \iint \left[\frac{1}{nh_2} \sum_{j=1}^{n} L\left(\frac{y-Y_j}{h_2}\right) \int \phi\left(x - Z_j, 0, \sqrt{h_1^2 + \sigma_u^2}\right) f_u(x-z)dx\right] f_{Z,Y}(z,y)dzdy$$

$$= \frac{1}{nh_2} \sum_{j=1}^{n} \iint L\left(\frac{y-Y_j}{h_2}\right) \phi\left(z - Z_j, 0, \sqrt{h_1^2 + 2\sigma_u^2}\right) f_{Z,Y}(z,y)dzdy.$$

Applying the leave-one-out strategy, and recall that L is the standard normal density function, we obtain

$$\iint \hat{f}_{X,Y}(x,y) f_{X,Y}(x,y)dxdy = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i} \phi(Y_i - Y_j, 0, h_2) \phi\left(Z_i - Z_j, 0, \sqrt{h_1^2 + 2\sigma_u^2}\right).$$

Therefore, the LSCV function to be minimized is

$$\begin{aligned} LSCV(h_1,h_2) =& \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(Y_i-Y_j,0,\sqrt{2}h_2)\phi\left(Z_i-Z_j,0,\sqrt{2(h_1^2+\sigma_u^2)}\right) \\ &-\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\phi(Y_i-Y_j,0,h_2)\phi\left(Z_i-Z_j,0,\sqrt{h_1^2+2\sigma_u^2}\right). \end{aligned} \tag{3.4.2}$$

3.5 Simulation

In this section, we conduct a simulation study to evaluate the performance of the LSCV bandwidth selection procedure. To generate the simulation data, we assume that $(Z,Y)^T$ follows a bivariate standard normal distribution. Y is directly observable, $X=Z+U$, and the measurement error U is distributed as a normal distribution with mean 0 and variance σ_u^2. Thus the latent variable X has a normal distribution with mean 0 and variance $1+\sigma_u^2$. The sample sizes are chosen to be $n=100,200$, and to see the effect of σ_u^2 on the LSCV results, we choose $\sigma_u^2=0.1$ and 0.5.

The performance of the estimation procedures is evaluated by the MSE. In specific, the estimator $\tilde{f}(x,y)$ and the true density function $f(x,y)$ are evaluated at equally spaced grid points from $[-3,3]\times[-3,3]$ in the step of 0.1. The MSE is then calculated as the average of the squared difference between the estimator and the true density values. For $\tilde{f}(x,y)$, the MSE based on a preselected bandwidth sequence $h=an^{-1/5}$ with $a=0.5,1,1.5,2,2.5,3,3.5$ and the bandwidth from the LSCV are calculated. The simulation is repeated 100 times, and the boxplots for each bandwidth selection are plotted in Figure 3.1 for $n=100$, where the xticks are the preselected bandwidths, and the last the boxplot in each panel is from LSCV. Clearly, the MSE based on the LSCV method performs satisfactorily, which is nearly the same as the lowest MSE produced by the preselected bandwidths. The patterns are the same for the both $\sigma_u^2=0.1$ and 0.5. The simulation results for $n=200$, as shown in Figure 3.2, again show that the LSCV bandwidth selection performs very well.

Figure 3.1: MSE boxplots of $\tilde{f}(x, y)$ when $n = 100$. The left plot is for $\sigma_u^2 = 0.1$, and the right plot is for $\sigma_u^2 = 0.5$

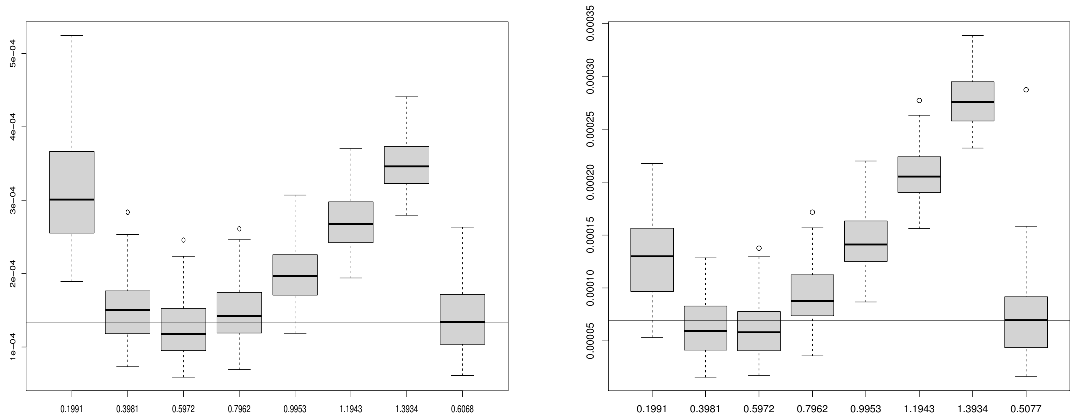


Figure 3.2: MSE boxplots of $\tilde{f}(x, y)$ when $n = 200$. The left plot is for $\sigma_u^2 = 0.1$, and the right plot is for $\sigma_u^2 = 0.5$

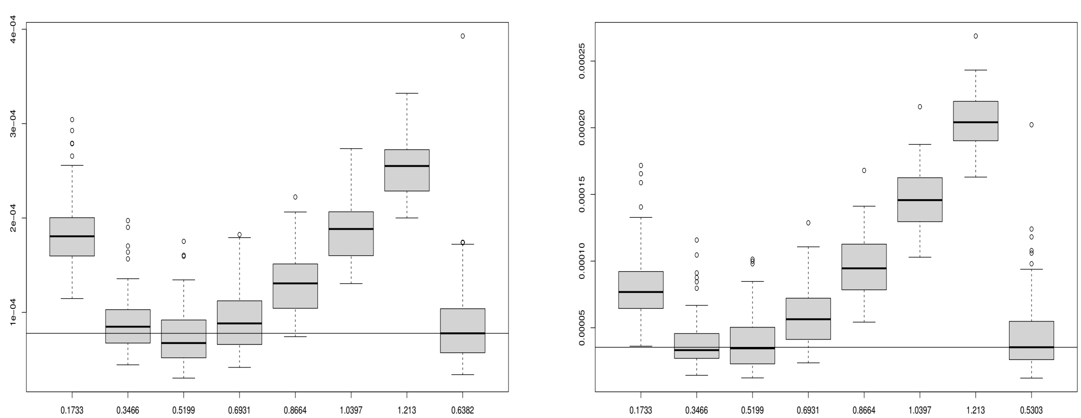
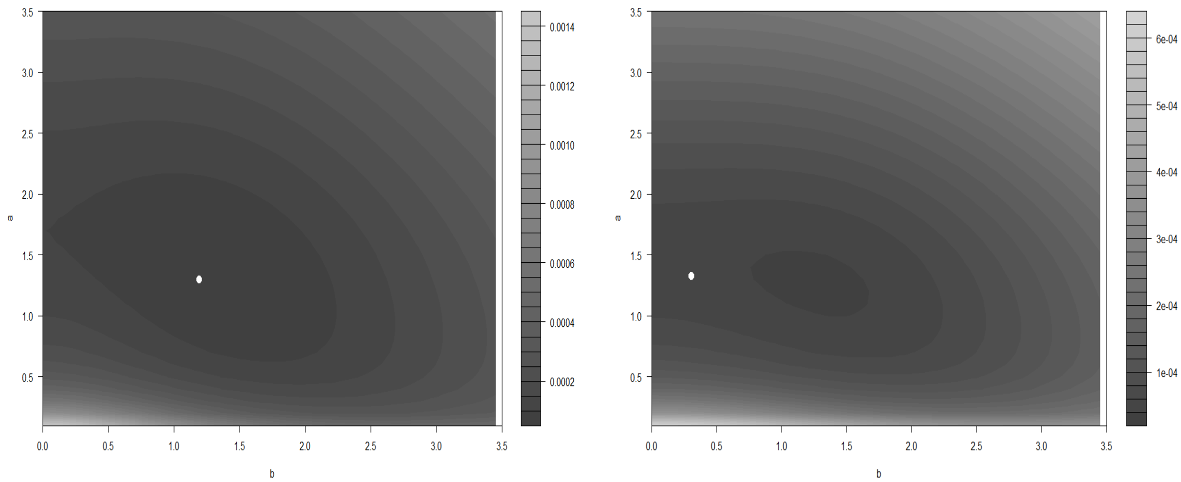


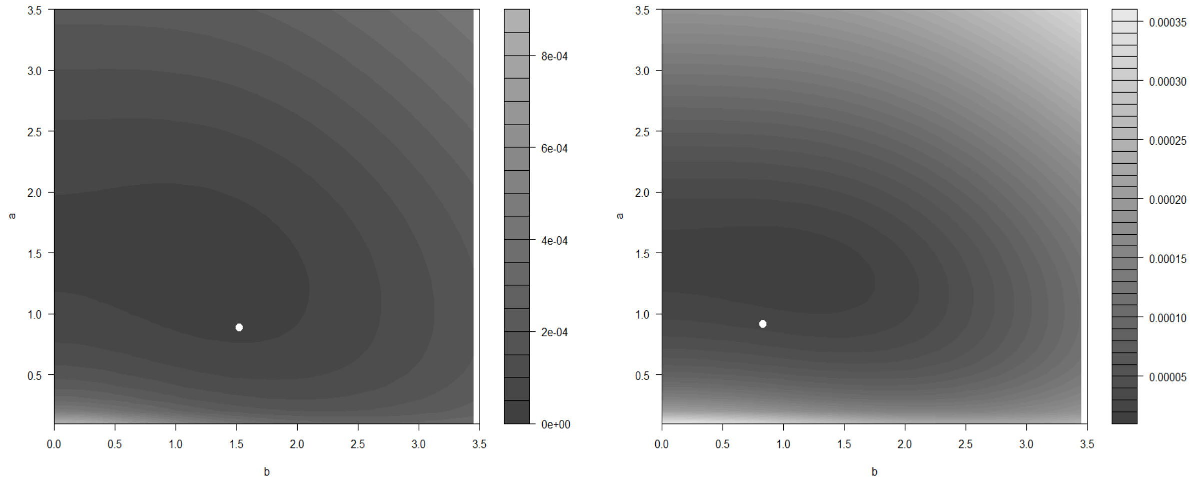
Figure 3.3: MSE boxplots of $\hat{f}(x, y)$ when $n = 100$. The left plot is for $\sigma_u^2 = 0.1$, and the right plot is for $\sigma_u^2 = 0.5$



For $\hat{f}(x, y)$, there are two bandwidths that we have to choose. To see the performance of the LSCV in this case, we choose $h_1 = an^{-1/5}$, $h_2 = bn^{-1/5}$, with a being equally selected from $[0.004, 3.5]$ in the step of 0.05, and b from $[0.1, 3.5]$ in the step of 0.1. Similar to $\tilde{f}(x, y)$, for each pair (h_1, h_2), the MSEs are calculated from the values of $\hat{f}(x, y)$ and the true density function $f(x, y)$ evaluated at equally spaced grid points from $[-3, 3] \times [-3, 3]$ in the step of 0.1. Figures 3.3 and 3.4 display the contour plot of the MSEs of $\hat{f}(x, y)$ based on different bandwidths, where the white dots represent the MSE estimated using the LSCV approach. It is evident that the LSCV technique performs well in all cases as the black dots are very close to the lowest MSEs.

Furthermore, one can see that when the variance of measurement error is small, the LSCV selection works better. However, when the measurement error variance increases to 0.5, the MSE obtained by LSCV deviates, although not very much, from the smallest MSE.

Figure 3.4: MSE boxplots of $\hat{f}(x, y)$ when $n = 200$. The left plot is for $\sigma_u^2 = 0.1$, and the right plot is for $\sigma_u^2 = 0.5$



3.6 Appendix

Proof of Theorem 3.3.1. Note that

$$
\begin{aligned}
E\tilde{f}_{X,Y}(x, y) &= \frac{1}{h} E f_u(x - Z_j) K\left(\frac{y - Y_j}{h}\right) = \frac{1}{h} \iint f_u(x - z) K\left(\frac{y - v}{h}\right) f_{Z,Y}(z, v) dz dv \\
&= \iint f_u(x - z) K(v) f_{Z,Y}(z, y - vh) dz dv,
\end{aligned}
$$

By (C2) and Taylor expansion, for each (z, y), we have

$$
f_{Z,Y}(z, y - vh) = f_{Z,Y}(z, y) - vh f'_{Z,Y}(z, y) + \frac{v^2 h^2}{2} f''_{Z,Y}(z, y) + o(h^2),
$$

where $f'Z, Y(z, y)$ is defined as the first derivative of $f_{Z,Y}(z, y)$ with respect to y. Thus,

$$
\begin{aligned}
E\tilde{f}_{X,Y}(x, y) &= \iint f_u(x - z) K(v) \left[f_u(z, y) - vh f'_{Z,Y}(z, y) + \frac{v^2 h^2}{2} f''_{Z,Y}(z, y) + o(h^2) \right] dz dv \\
&= \int f_u(x - z) f_{Z,Y}(z, y) dz + \frac{h^2}{2} \int f_u(x - z) f''_{Z,Y} dz \int v^2 K(v) dv + o(h^2).
\end{aligned}
$$

By (C1), we know the joint density function of (U, Z, Y) is $f_{Z,Y}(z, y) f_U(u)$. It is easy to see that the joint density function of (X, Z, Y) can be written as $f_{Z,Y}(z, y)f_u(x - z)$, hence, the joint density function of (X, Y) satisfies

$$f_{X,Y}(x, y) = \int f_{Z,Y}(z, y)f_U(x - z)dz. \tag{3.6.1}$$

This implies

$$E\tilde{f}_{X,Y}(x, y) = f_{X,Y}(x, y) + \frac{h^2}{2}\int f_u(x - z)f''_{Z,Y}(z, y)dy + o(h^2).$$

To calculate the variance of $\tilde{f}_{X,Y}(x, y)$, we know that

$$\begin{aligned}
&Ef_u^2(x - z)K^2\left(\frac{y - Y}{h}\right) = \int f_U^2(x - z)K^2\left(\frac{y - v}{h}\right)f_{Z,Y}(z, v)dzdv \\
&= h\int f_u^2(x - z)K^2(v)f_{Z,Y}(z, y - vh)dzdv \\
&= h\int f_U^2(x - z)K^2(v)\left[f_{Z,Y}(z, y) - vhf'_{Z,Y}(z, y) + \frac{v^2h^2}{2}f''_{Z,Y}(z, y) + (h^2)\right]dzdv \\
&= hR(K)\int f_U^2(x - z)f_{Z,Y}(z, y)dz + o(h^2).
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathrm{Var}(\tilde{f}_{X,Y}(x, y)) = \frac{1}{nh^2}\mathrm{Var}\left(f_u(x - z)K\left(\frac{y - Y}{h}\right)\right) \\
&= \frac{1}{nh^2}\left[h\int f_U^2(x - z)f_{Z,Y}(x, z)dz \cdot R(K) + o(h^2)\right] - \frac{1}{n}(E\hat{f}_{X,Y}(x, y))^2 \\
&= \frac{1}{nh^2}\left[h\int f_U^2(x - z)f_{Z,Y}(z, y)dz \cdot R(K) + O(h^3)\right] - \frac{1}{n}\left[f_{X,Y}(x, y) + O(h^2)\right]^2 \\
&= \frac{1}{nh}\int f_U^2(x - z)f_{Z,Y}(z, y)dz \cdot R(K) + O\left(\frac{1}{nh}\right).
\end{aligned}$$

So, the MSE of $\tilde{f}_{X,Y}(x, y)$ can be derived from the bias-variance decomposition and the expectation and variance expressions derived above. □

Proof of Theorem 3.3.2. Denote that

$$\xi_{jn} = \frac{1}{nh} f_U(x - Z_j) K\left(\frac{y - Y_j}{h}\right) - E\frac{1}{nh} f_U(x - Z_j) K\left(\frac{y - Y_j}{h}\right).$$

Then $\tilde{f}_{X,Y}(x, y) - E\tilde{f}_{X,Y}(x, y) = \sum_{j=1}^{n} \xi_{jn}$. We will check this independent sum satisfies the Lyapunov-Condition. Clearly, $E(\xi_{jn}) = 0$. For any $\gamma > 0$, by C_r-inequality,

$$E\left|\xi_{jn}\right|^{2+\gamma} \leq \frac{c}{(nh)^{2+\gamma}} E\left[f_U(x - Z_j) K\left(\frac{y - Y_j}{h}\right)\right]^{2+\gamma} + \frac{C}{n^{2+\gamma}}\left[E\frac{1}{h} f_U(x - z) K\left(\frac{y - Y}{h}\right)\right]^{2+\gamma}.$$

Note that

$$\begin{aligned}
&E\left[f_U(x - Z_j) K\left(\frac{y - Y_j}{h}\right)\right]^{2+\gamma} = \iint f_U^{2+\gamma} K^{2+\gamma}\left(\frac{y - v}{h}\right) f_{Z,Y}(z, y) dz dv \\
&= h \iint f_U^{2+\gamma}(x - z) K^{2+\gamma}(v) f_{Z,Y}(z, y - vh) dz dv \\
&= h \iint f_U^{2+\gamma}(x - z) K^{2+\gamma}(v)[f_{Z,Y}(z, y) - vh f'_{Z,Y}(z, y) + \frac{v^2 h^2}{2} f''_{Z,Y}(z, y) + o(h^2)] dz dv = O(h),
\end{aligned}$$

so, $E|\xi_{jn}|^{2+\gamma} = (nh)^{-2-\gamma} O(h) + n^{-2-\gamma} O(1) = O(n^{-2-\gamma} h^{-1-\gamma})$. From the variance deviation we also know that $E\xi_{jn}^2 = O(\frac{1}{n^2 h})$. Therefore,

$$\frac{nE|\xi_{jn}|^{2+\gamma}}{(nE\xi_{jn}^2)^{\frac{2+\gamma}{2}}} = \frac{O\left(n \cdot \frac{1}{n^{2+\gamma} h^{1+\gamma}}\right)}{O\left((n \cdot \frac{1}{n^2 h})^{\frac{2+\gamma}{2}}\right)} = O\left(\frac{1}{(nh)^{\frac{\gamma}{2}}}\right)$$

which is the order of $o(1)$ by (C4). Application of the Lyapunov-CLT finishes the proof of Theorem 3.3.2. □

Proof of Theorem 3.3.3. We only have to show that $\tilde{f}_{X,Y}(x, y) - E\tilde{f}_{X,Y}(x, y) \to 0$ almost surely. Redefine

$$\xi_{nj} = f_U(x - Z_j) K\left(\frac{y - Y_j}{h}\right) - E f_U(x - Z) K\left(\frac{y - Y}{h}\right).$$

Then $h[\tilde{f}_{X,Y}(x, y) - E\tilde{f}_{X,Y}(x, y)] = n^{-1} \sum_{j=1}^{n} \xi_{nj}$. Note that ξ_{nj} are i.i.d. with $E\xi_{nj} = 0$ and

$|\xi_{nj}| \le 2\sup_u K(u) := B$ by the boundedness of K, as assumed in (C3). From the proof of Theorem 3.3.1, we know that

$$\sigma_n^2 := \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(\xi_{ni}) = nh^2\mathrm{Var}(\tilde{f}_{X,Y}(x,y)) = hR(K)\int f_U^2(x-z)f_{Z,Y}(z,y)dz + o(h).$$

From Bernstein-inequality, for $\forall\epsilon > 0$ and n sufficiently large,

$$P(|\tilde{f}_{X,Y}(x,y) - Ef_{X,Y}(x,y)| \ge \epsilon) = P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{ni}\right| \ge \epsilon h\right) \le 2\exp\left(-\frac{n\epsilon^2h^2}{2\sigma_n^2 + B\epsilon h}\right)$$
$$\le 2\exp\left(-\frac{\epsilon^2 nh}{2R(K)\int f_U^2(x-z)f_{Z,Y}z,y)dz + o(1) + B\epsilon}\right).$$

Note that $nh/\log n \to \infty$ by assumption (C6), so for some positive constant c,

$$\sum_{n=1}^{\infty} P(|\tilde{f}_{X,Y}(x,y) - E\tilde{f}_{X,Y}(x,y)| \ge \epsilon) \le c\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty,$$

which implies $\tilde{f}_{X,Y} - E\tilde{f}_{X,Y} = o(1)$ almost surely by Borel-Cantelli Lemma. □

Proof of Theorem 3.3.4. Note that

$$\begin{aligned}
& \left|E\tilde{f}_{X,Y}(x,y) - f_{X,Y}(x,y)\right| \\
= & \left|\frac{1}{n}\iint f_U(x-z)K\left(\frac{y-v}{h}\right)f_{Z,Y}(z,y)dzdv - f_{X,Y}(x,y)\right| \\
= & \left|\frac{1}{h}\iint f_U(x-z)K\left(\frac{y-v}{h}\right)f_{Z,Y}(z,y)dzdv - \int f_U(x-z)f_{Z,Y}(z,y)dz\right| \\
= & \left|\iint \frac{1}{h}f_U(x-z)K\left(\frac{y-v}{h}\right)[f_{Z,Y}(z,v) - f_{Z,Y}(z-y)]dzdv\right| \\
= & \left|\iint \frac{1}{h}f_U(x-z)K(\frac{v}{h})[f_{Z,Y}(z,y-v) - f_{Z,Y}(z-y)]dzdv\right| \\
= & \left|\int \frac{1}{h}K\left(\frac{v}{h}\right)\left[\int f_U(x-z)f_{Z,Y}(z,y-v)dz - \int f_U(x-z)f_{Z,Y}(z,y)dz\right]dv\right| \\
= & \left|\int \frac{1}{n}K\left(\frac{v}{h}\right)[f_{X,Y}(x,y-v) - f_{X,Y}(x,y)]\,dv\right| \\
\le & \sup_{x,y}\cdot\sup_{|u|<\delta}|f_{X,Y}(x,y-v) - f_{X,Y}(x,y)| + 2B\cdot\int_{|u|>\frac{\delta}{h}}K(u)du,
\end{aligned}$$

which converges to 0 as $n \to \infty$. Note that $K(v) = (2\pi)^{-1} \int e^{-ivt}\phi_K(t)dt$, then

$$\tilde{f}_{X,Y}(x,y) = \frac{1}{nh}\sum_{j=1}^{n} f_U(x-Z_j)K\left(\frac{y-Y_j}{h}\right) = \frac{1}{n}\sum_{j=1}^{n} f_U(x-Z_j)\frac{1}{2\pi h}\int e^{-i\left(\frac{y-Y_j}{h}\right)t}\phi_K(t)dt$$

$$= \frac{1}{2n\pi}\sum_{j=1}^{n} f_U(x-Z_j)\int e^{-i(y-Y_j)t}\phi_K(th)dt = \frac{1}{2n\pi}\int e^{-iyt} f_U(x-Z_j)e^{iY_jt})\phi_K(th)dt.$$

For the characteristic function $\phi_U(t)$ of U, we have $f_U(x-Z_j) = (2\pi)^{-1}\int e^{-is(x-Z_j)}\phi_u(s)ds$. Thus,

$$\tilde{f}_{X,Y}(x,y) = \frac{1}{2\pi}\iint e^{-iyt}e^{-isx}\left(\frac{1}{n}\sum_{j=1}^{n} e^{iY_jt+iZ_js}\right)\phi_U(s)\phi_K(th)dtds.$$

This implies that $E\sup_{x,y}|\tilde{f}_{X,Y}(x,y) - E\tilde{f}_{X,Y}(x,y)|$ is bounded above by

$$E\iint |\phi_U(s)\phi_K(th)|\cdot|\phi_n(t,s)-E\phi_n(t,s)|dtds = \iint \phi_U(s)\phi_K(th)\cdot E|\phi_n(t,s)-E\phi_n(t,s)|dtds,$$

where $\phi_n(t,s) = n^{-1}\sum_{j=1}^{n} e^{iY_jt+iZ_js}$. Note that $E|\phi_n(t,s) - E\phi_n(t,s)|^2 \le n^{-1}$. Then

$$E\sup_{x,y}|\tilde{f}_{X,Y}(x,y)-E\tilde{f}_{X,Y}(x,y)| \le \frac{1}{\sqrt{n}}\iint |\phi_U(s)\phi_K(th)|dsdt = \frac{1}{\sqrt{nh^2}}\iint |\phi_U(s)\phi_K(t)|dsdt,$$

which converges to 0 since $nh^2 \to \infty$ by (C9). Thus, the result of Theorem 3.3.4 can be obtained by the following triangular inequality

$$\sup_{x,y}|\tilde{f}_{X,Y}(x,y) - f_{X,Y}(x,y)| \le \sup_{x,y}|\tilde{f}_{X,Y}(x,y) - E\tilde{f}_{X,Y}(x,y)| + \sup_{x,y}|E\tilde{f}_{X,Y}(x,y) - f_{X,Y}(x,y)|$$

and the results derived above. □

Proof of Theorem 3.3.5. First, we calculate the expectation of $\hat{f}_{X,Y}(x,y)$.

$$\begin{aligned}
E\hat{f}_{X,Y}(x,y) &= \frac{1}{h_1h_2}E\left[\int K\left(\frac{x-u-Z}{h_1}\right)f_u(u)du\cdot L\left(\frac{y-Y}{h_2}\right)\right] \\
&= \frac{1}{h_1h_2}\iint\left[\int K\left(\frac{x-u-z}{h_1}\right)f_U(u)du\right]\cdot L\left(\frac{y-v}{h_2}\right)f_{Z,Y}(z,v)dzdv \\
&= \frac{1}{h_1h_2}\iiint K\left(\frac{x-u-z}{h_1}\right)L\left(\frac{y-v}{h_2}\right)f_U(u)f_{Z,Y}(z,v)dudzdv \\
&= \iiint K(z)L(v)f_U(u)f_{Z,Y}(x-u-zh_1,y-vh_2)dudzdv \\
&= \iint K(z)L(v)f_{X,Y}(x-zh_1,y-vh_2)dzdv.
\end{aligned}$$

The last inequality is based on (3.6.1). By Taylor expansion, we have

$$\begin{aligned}
f_{X,Y}(x-zh_1,y-vh_2) = f_{X,Y}(x,y) &- zh_1\frac{\partial f_{X,Y}(x,y)}{\partial x} - vh_2\frac{\partial f_{X,Y}(x,y)}{\partial y} + \frac{z^2h_1^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial x^2} \\
&+ zvh_1h_2\frac{\partial f_{X,Y}(x,y)}{\partial x\partial y} + \frac{v^2h_2^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial y^2} + o(h_1^2) + o(h_2^2) + o(h_1h_2).
\end{aligned}$$

So, $E\hat{f}_{X,Y}(x,y)$ can be written as

$$\begin{aligned}
&\iint K(z)L(v)\left[f_{X,Y}(x,y) - zh_1\frac{\partial f_{X,Y}(x,y)}{\partial x} - vh_2\frac{\partial f_{X,Y}(x,y)}{\partial y} + \frac{z^2h_1^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial x^2}\right. \\
&\qquad \left.+zvh_1h_2\frac{\partial f_{X,Y}(x,y)}{\partial x\partial y} + \frac{v^2h_2^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial y^2} + o(h_1^2) + o(h_2^2) + o(h_1h_2)\right]dzdv \\
&= f_{X,Y}(x,y) + \frac{h_1^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial x^2} + \frac{h_2^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial y^2} + o(h_1^2) + o(h_2^2) + o(h_1h_2).
\end{aligned}$$

Next, we consider the variance of $\hat{f}_{X,Y}(x,y)$. Note that

$$\mathrm{Var}(\hat{f}_{X,Y}(x,y)) = \frac{1}{nh_1^2h_2^2}\mathrm{Var}\left(L\left(\frac{y-Y_j}{h_2}\right)\int K\left(\frac{x-u-z}{h_1}\right)f_U(u)du\right)$$

and

$$
\begin{aligned}
& E\left[L\left(\frac{y-Y}{h_2}\right)\int K\left(\frac{x-u-z}{h_1}\right)f_U(u)du\right]^2 \\
= & \iint\left[\int K\left(\frac{x-u-z}{h_1}\right)f_U(u)du\right]^2 L^2\left(\frac{y-v}{h_2}\right)f_{Z,Y}(z,v)dzdv \\
= & \iint\left[h_1\int K(u)f_U(x-z-uh_1)du\right]^2 h_2L^2(v)f_{Z,Y}(z,y-vh_2)dzdv \\
= & \iint\left[h_1\int K(u)[f_U(x-z)-uh_1f_U'(x-z)+\frac{u^2h_1^2}{2}f_U''(x-z)+O(h_1^3)]du\right]^2\cdot \\
& \qquad h_2L^2(v)f_{Z,Y}(z,y-vh_2)dzdv \\
= & \iint\left[h_1\int K(u)f_U(x-z)du-h_1^2\int uK(u)f_U'(x-z)du+O(h_1^3)\right]^2\cdot \\
& \qquad h_2L^2(v)\left[f_{Z,Y}(z,y)-vh_2f_{Z,Y}'(z,y)+\frac{v^2h_2^2}{2}f_{Z,Y}''(z,y)+O(h_2^3)\right]dzdv \\
= & \iint[h_1f_U(x-z)+O(h_1^2)]^2\cdot[h_2L^2(v)f_{Z,Y}(z,y)+O(h_2^3)]dzdv \\
= & \ h_1^2h_2R(L)\int f_U^2(x-z)f_{Z,Y}(z,y)dz+o(h_1^2h_2).
\end{aligned}
$$

Hence, $\mathrm{Var}(\hat{f}_{X,Y}(x,y))$ can be written as

$$
\begin{aligned}
& \frac{1}{nh_1h_2}\left[h_1^2h_2R(L)\int f_U^2(x-z)f_{Z,Y}(z,y)dz+O(h_1^2h_2^3)+O(h_1^4h_2)\right] \\
& -\frac{1}{nh_1^2h_2}\cdot\left[h_1h_2(f_{X,Y}(x,y)+\frac{h_1^2}{2}\frac{\partial f_{X,Y}^2(x,y)}{\partial x^2}+\frac{h_2^2}{2}\cdot\frac{\partial f_{X,Y}^2(x,y)}{\partial y^2}+O(h_1^3)+O(h_2^3))\right]^2 \\
= & \frac{R(L)}{nh_2}\int f_U^2(x-z)f_{Z,Y}(z,y)dz+O\left(\frac{h_2}{n}\right)+O\left(\frac{h_1^2}{nh_2}\right)-\frac{1}{n}\left[f_{X,Y}(x,y)+O(h_1^2)+O(h_2^2)\right]^2 \\
= & \frac{R(L)}{nh_2}\int f_U^2(x-z)f_{Z,Y}(z,y)dz+O\left(\frac{h_2}{n}\right)+O\left(\frac{h_1^2}{nh_2}\right)+O\left(\frac{1}{n}\right).
\end{aligned}
$$

The expression of MSE then can be obtained by the routine bias and variance decomposition formula. □

Proof of Theorem 3.3.6. Define

$$\xi_{jn} = \frac{1}{nh_1h_2}\int K\left(\frac{x-u-Z_j}{h_1}\right) f_U(u)du \cdot L\left(\frac{y-Y_j}{h_2}\right) \\ - E\frac{1}{nh_1h_2}\int K\left(\frac{x-u-Z_j}{h_1}\right) f_U(u)du \cdot L\left(\frac{y-Y_j}{h_2}\right).$$

Clearly, $E\xi_{jn} = 0$. there exists a positive constant c such that

$$E|\xi_{jn}|^{2+\delta} \leq \frac{c}{(nh_1h_2)^{2+\delta}} E\left[\int K\left(\frac{x-u-Z_j}{h_1}\right) f_U(u)du \cdot L\left(\frac{y-Y_j}{h_2}\right)\right]^{2+\delta} \\ + \frac{c}{n^{2+\delta}}\left[E\frac{1}{h_1h_2}\int K\left(\frac{x-u-Z_j}{h_1}\right) f_U(u)du \cdot L\left(\frac{y-Y_j}{h_2}\right)\right]^{2+\delta}.$$

Note that

$$E\left[\int K\left(\frac{x-u-z}{h_1}\right) f_U(u)du L\left(\frac{y-Y}{h_2}\right)\right]^{2+\delta} \\ = \iint \left[\int K\left(\frac{x-u-z}{h_1}\right) f_U(u)du\right]^{2+\delta} L^{2+\delta}\left(\frac{y-v}{h_2}\right) f_{Z,Y}(z,v)dzdv \\ = h_1^{2+\delta}h_2 \iint \left[\int K(u)f_U(x-z-uh_1)du\right]^{2+\delta} L^{2+\delta}(v) f_{Z,Y}(z, y-vh_2)dzdv,$$

which is the order of $O(h_1^{2+\delta}h_2)$ by (C5). So,

$$E\xi_{jn}^{2+\delta} = \frac{1}{(nh_1h_2)^{2+\delta}}O(h_1^{2+\delta}h_2) + \frac{1}{n^{2+\delta}}O(1) = O\left(\frac{1}{n^{2+\delta}h_2^{1+\delta}}\right).$$

Note that $\hat{f}_{X,Y}(x,y) - E\hat{f}_{X,Y}(x,y) = \sum_{j=1}^{n}\xi_{jn}$ and $\mathrm{Var}(\hat{f}_{X,Y}(x,y)) = nE(\xi_{jn}^2)$. From the variance calculation in the proof of Theorem 3.3.5 we know that $E(\xi_{jn}^2) = O(n^{-2}h_2^{-1})$. Thus, we have

$$\frac{nE|\xi_{jn}|^{2+\delta}}{\left(nE\xi_{jn}^2\right)^{\frac{2+\delta}{2}}} = \frac{O\left(\frac{n}{n^{2+\delta}h_2^{1+\delta}}\right)}{O\left(\left(\frac{n}{n^2h_2}\right)^{\frac{2+\delta}{2}}\right)} = O\left(\frac{1}{(nh_2)^{\frac{\delta}{2}}}\right),$$

which is the order of $o(1)$. Thus the conclusion in Theorem 3.3.6 can be drawn by Lyapunov-

CLT. □

Proof of Theorem 3.3.7. Rewrite $h_1h_2[\hat{f}_{X,Y}(x,y) - E\hat{f}_{X,Y}(x,y)]$ as $n^{-1}\sum_{j=1}^n \xi_{nj}$, where

$$\xi_{nj} = L\left(\frac{y-Y_j}{h_2}\right)\int K\left(\frac{x-u-Z_j}{h_1}\right)f_U(u)du - EL\left(\frac{y-Y}{h_2}\right)\int K\left(\frac{x-u-Z}{h_1}\right)f_U(u)du.$$

ξ_{nj} are i.i.d. with $E\xi_{nj}=0$ and $|\xi_{nj}| \le 2\cdot\sup_u K(u)L(u) = B$ by the boundedness of K and L. We assume that f_U, K and L are bounded. Then we have

$$\sigma_n^2 = nh_1^2h_2^2\mathrm{Var}(\hat{f}_{X,Y}(x,y)) = h_1^2h_2\int f_U^2(x-z)f_{Z,Y}(z,y)dz\int L^2(v)dv + o(h_1^2h_2).$$

From Bernstein-inequality, for $\forall\epsilon>0$, and n sufficiently large,

$$\begin{aligned}
&P\left(|\hat{f}_{X,Y}(x,y) - E\hat{f}_{X,Y}(x,y)| \ge \epsilon\right)\\
&= P\left(\left|\frac{1}{n}\sum_{j=1}^n \xi_{nj}\right| \ge \epsilon h_1h_2\right) \le 2\exp\left(-\frac{n(\epsilon h_1h_2)^2}{2\sigma_n^2 + B\epsilon h_1h_2}\right)\\
&= 2\exp\left(-\frac{\epsilon^2 nh_1h_2}{2h_1\int f_U^2(x-z)f_{Z,Y}(z,y)dz\int L^2(v)dv + o(1) + B\epsilon}\right).
\end{aligned}$$

By assumption (D4), $nh_1h_2/\log n \to \infty$, then for a positive constant c.

$$\sum_{n=1}^\infty P(|\hat{f}_{X,Y}(x,y) - E\hat{f}_{X,Y}(x,y)| \ge \epsilon) \le c\sum_{n=1}^\infty \frac{1}{n^2} < \infty.$$

By Borel-Cantelli lemma, we finish the proof of Theorem 3.3.7. □

Proof of Theorem 3.3.8. Note that $\left|E\hat{f}_{X,Y}(x,y) - f_{X,Y}(x,y)\right|$ is equal to

$$
\begin{aligned}
&\left|\frac{1}{h_1 h_2}\iint \left[\int K\left(\frac{x-u-z}{h_1}\right) f_U(u)du\right] L\left(\frac{y-v}{h_2}\right) f_{Z,Y}(z,v)dzdv - f_{X,Y}(x,y)\right| \\
&= \left|\frac{1}{h_1 h_2}\iint K\left(\frac{z}{h_1}\right) L\left(\frac{v}{h_2}\right) [f_{X,Y}(x-z,y-v) - f_{X,Y}(x,y)]dzdv\right| \\
&\le \sup_{x,y}\sup_{|v|<\delta_1}\sup_{|z|<\delta_2} |f_{X,Y}(x-z,y-v) - f_{X,Y}(x,y)| \int \frac{1}{h_1}K\left(\frac{z}{h_1}\right)dz \int \frac{1}{h_2}L\left(\frac{v}{h_2}\right)dv \\
&\qquad + 2B\Big(\int_{|u_1|>\frac{\delta_1}{h_1}} K(u_1)du_1 \int_{|u_2|>\frac{\delta_2}{h_2}} L(u_2)du_2 + \int_{|u_1|>\frac{\delta_1}{h_1}} K(u_1)du_1 \int_{|u_2|<\frac{\delta_2}{h_2}} L(u_2)du_2 \\
&\qquad + \int_{|u_1|<\frac{\delta_1}{h_1}} K(u_1)du_1 \int_{|u_2|>\frac{\delta_2}{h_2}} L(u_2)du_2\Big),
\end{aligned}
$$

so it converges to 0 as $n \to \infty$. Rewrite K and L as $K(v) = (2\pi)^{-1}\int e^{-ivt_1}\phi_K(t_1)dt_1$ and $L(v) = (2\pi)^{-1}\int e^{-ivt_2}\phi_L(t_2)dt_2$, then we have

$$
\begin{aligned}
&\hat{f}_{X,Y}(x,y) = \frac{1}{n}\sum_{j=1}^{n}\int \frac{1}{h_1}K\left(\frac{x-u-Z_j}{h_1}\right) f_U(u)du \cdot \frac{1}{h_2}L\left(\frac{y-Y_j}{h_2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\int \left[\frac{1}{2\pi h_1}\int e^{-i\left(\frac{x-u-Z_j}{h_1}\right)t_1}dt_1\right] f_U(u)du \frac{1}{2\pi h_2}\int e^{-i\left(\frac{y-Y_j}{h_2}\right)t_2}\phi_L(t_2)dt_2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{2\pi}\right)^2 \iint e^{-i(x-u-Z_j)t_1}\phi_K(th_1)f_U(u)dt_1du \int e^{-i(y-Y_j)t_2}\phi_L(t_2h_2)dt_2 \\
&= \left(\frac{1}{2\pi}\right)^2 \iiint e^{-i(x-u)t_1}e^{-iyt_2}\left(\frac{1}{n}\sum_{j=1}^{n} e^{iZ_jt_1}e^{iY_jt_2}\right)\phi_K(t_1h_1)\phi_L(t_2h_2)f_U(u)dudt_1dt_2.
\end{aligned}
$$

Note that $\phi_U(t_1) = \int e^{it_1u}f_U(u)du$, so

$$
\hat{f}_{X,Y}(x,y) = \left(\frac{1}{2\pi}\right)^2 \iint e^{-ixt_1}e^{-iyt_2}\cdot\left(\frac{1}{n}\sum_{j=1}^{n} e^{iZ_jt_1}e^{iY_jt_2}\right)\phi_K(t_1h_1)\phi_L(t_2h_2)\phi_U(t_1)dt_1dt_2.
$$

This implies that the expectation $E\sup_{x,y}|\hat{f}_{X,Y}(x,y)-E\hat{f}_{X,Y}(x,y)|$ is bounded above by

$$\begin{aligned}
&E\iint |\phi_K(t_1h_1)\phi_L(t_2h_2)\phi_U(t_1)|\cdot|\phi_n(t_1,t_2)-E\phi_n(t_1,t_2)|dt_1dt_2\\
=&\iint |\phi_K(t_1h_1)\phi_L(t_2h_2)\phi_U(t_1)|\cdot E|\phi_n(t_1,t_2)-E\phi_n(t_1,t_2)|dt_1dt_2,
\end{aligned}$$

where $\phi_n(t_1,t_2)=n^{-1}\sum_{j=1}^n e^{iY_jt_1+iZ_jt_2}$. Note that $E[\phi_n(t_1,t_2)-E\phi_n(t_1,t_2)]^2\le n^{-1}$. Then

$$\begin{aligned}
E\sup_{x,y}|\hat{f}_{X,Y}(x,y)-E\hat{f}_{X,Y}(x,y)| &\le \frac{1}{\sqrt{n}}\int|\phi_U(t_1)|dt_1\int|\phi_L(t_2h_2)|dt_2\\
&= \frac{1}{\sqrt{nh_2^2}}\int|\phi_U(t_1)|dt_1\phi_K(t_1h_1)|dt_1\int|\phi_L(t_2)|dt_2,
\end{aligned}$$

which converges to 0 since $nh_2^2\to\infty$. Thus, applying the triangular inequality

$$\sup_{x,y}|\hat{f}_{X,Y}(x,y)-f_{X,Y}(x,y)|\le\sup_{x,y}|\hat{f}_{X,Y}(x,y)-E\hat{f}_{X,Y}(x,y)|+\sup_{x,y}|E\hat{f}_{X,Y}(x,y)-f_{X,Y}(x,y)|$$

leads to the conclusion of the theorem. $\square$

Chapter 4

Conclusion

In this dissertation, several nonparametric density estimators are proposed for the data contaminated super-smooth, ordinary-smooth, Berkson measurement errors. The estimators are built using classical kernel and deconvolution kernel smoothing as building blocks.

In chapter 2, we proposed a nonparametric mixed kernel estimator for a multivariate density function and its derivatives when the data are contaminated with different sources of measurement errors. This estimator combines classical and deconvolution kernels to account for error-free and error-prone variables, respectively, which is an extension of the estimator in Shi et al. (2020). In addition, we consider the case where all three types of variables are presented in the data, also, each type of the variable is multidimensional.

Although Shi et al. (2020) showed that for the bivariate nonparametric density estimator of an error-free variable and a variable contaminated with measurement error, the optimal convergence rate is dominated by the super smooth or the ordinary smooth component. It might be natural to expect the optimal convergence rate is controlled in turn by the super smooth, ordinary smooth and the error-free data in the multivariate cases, but it is not clear how the optimal convergence rate is determined within each type of measurement errors. In chapter 2, we identified which measurement error type, and which variable within each measurement error type, plays the dominated role in determining the optimal convergence rate of nonparametric estimators of the joint density or its derivatives.

Further more, we discussed about the bandwith selection issue. As we showed in this chapter 2, Theorem 2.4.4 shows that the optimal convergence rate can not be improved by choosing different bandwidths for different components; when the super-smooth measurement error appears in the data, the optimal convergence rate depends on the order of derivatives, not on the dimensions of the variables. However, when only the ordinary-smooth measurement error and error-free variables appear in the data, the optimal convergence rate not only relies on the order of the derivatives, but also depends on the dimensions of the variables explicitly.

Another significance of this chapter 2 is that we established a minimax optimal convergence rate among ALL nonparametric estimators when the target joint density function or the derivatives of joint density function of the random triple lies in a smooth family, and proved that the proposed mixed kernel density estimator achieves the minimax optimal convergence rate.

In chapter 3, we proposed two estimators for the joint density function of two random variables, when one variable is contaminated with Berkson measurement error and another variable can be observed directly. Those two estimators were developed with or without applying the kernel smoothing for the data with Berkson measurement error.

The first estimator $\tilde{f}$ is based on the estimator proposed by Delaigle (2007), which is a combination of this component with the classical kernel smoothing idea for the observed data. Due to the drawback noted in Long et al. (2016), we proposed the second estimator by smoothing the Berkson measurement error component. For these two estimators, we have shown some theoretical comparisons, including their mean squared errors, weak and strong consistencies, and asymptotic normalities.

Furthermore, we explored the bandwidth selection issue, which has a substantial impact on a kernel density estimator's performance. Because the optimal bandwidth determined in section 3.3 has limitations, we developed a data-driven bandwidth selector for both estimators named least squares cross validation bandwidth selector. This selector was also shown to have a robust performance in terms of selecting the ideal bandwidth in order to minimize the estimators' mean squares error.

Though the probability density function are frequently the focus of attention, their estimation also serves as a crucial building block for other objects being modeled, such as a conditional mean (i.e., a ”regression function”), which can be directly modeled using nonparametric or semiparametric methods (a conditional mean is a function of a conditional PDF, which is itself a ratio of unconditional PDFs). Therefore, in the future work, we are going to develop a nonparametric regression method for these types of data.

Bibliography

M Rosenblatt. Remarks on some nonparametric estimates of a density function. annals of mathematical statistics. 1956.

Emanuel Parzen. On estimation of a probability density function and mode. The annals of mathematical statistics, 33(3):1065–1076

Wayne A Fuller. Measurement error models. John Wiley & Sons, 2009.

Amy F Subar, Frances E Thompson, Victor Kipnis, Douglas Midthune, Paul Hurwitz, Suzanne McNutt, Anna McIntosh, and Simon Rosenfeld. Comparative validation of the block, willett, and national cancer institute food frequency questionnaires: the eating at america's table study. American journal of epidemiology, 154(12):1089–1099

Vernon R Young, Luis Fajardo, Edwina Murray, William M Rand, and Nevin S Scrimshaw. Protein requirements of man: Comparative nitrogen balance response within the submaintenance-to-maintenance range of intakes of wheat and beef proteins. The Journal of nutrition, 105(5):534–542

Raymond J Carroll, David Ruppert, and Leonard A Stefanski. Measurement error in nonlinear models, volume 105. CRC press, 1995.

R. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. Journal of American Statistical Association, 83:1184–1186, 1988.

L. Stefanski and R. Carroll. Deconvolving kernel density estimators. Statistics, 21:169–184, 1990.

M.P. Wand and M.C. Jones. Kernel smoothing. Chapman & Hall/CRC, 1995.

J. Fan and I. Gijbels. Local polynomial modelling and its applications. Chapman & Hall/CRC, 1996.

Y.D. Wang. Smoothing splines: methods and application. CRC Press, 2011.

J.Q. Fan. on the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist., 19(3):1257–1272, 1991a.

J.Q. Fan. asymptotic noramlity for deconvolution kernel density estimators. Sankhyā, 53: 97–110, 1991b.

Jianqing Fan and Elias Masry. Multivariate regression estimation with errors-in-variables: Asymptotic normality for mixing processes. Journal of Multivariate Analysis, 43(2):237–271, 1992.

E. Masry. Multivariate probability density deconvolution for stationary random processes. IEEE Trans. Inform. Theory, 37:1105–1115, 1991.

E. Masry. asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. Journal of Multivariate Analysis, 44(1):47–68, 1993a.

Elias Masry. strong consistency and rates for deconvolution of multivariate densities of stationary processes. Stochastic Processes and their Applications, 47(1):53–74, 1993b.

Elias Masry. multivariate regression estimation with errors-in-variables for stationary processes. Journal of Nonparametric Statistics, 3(1):13–36, 1993c.

X.L. Tang. Convergence rates of multivariate deconvolution and compound Poisson estimation. PhD thesis, Purdue University, 1994.

J.H. Shi, X.Q. Bai, and W.X. Song. Partial deconvolution estimation in nonparametric regression. Canadian Journal of Statistics-Revue Canadienne De Statistique, 48(3):535–560, 2020.

A. Delaigle. Nonparametric density estimation from data with a mixture of berkson and classical errors. Canadian Journal of Statistics-Revue Canadienne De Statistique, 35(1): 89–104, 2007.

R. J. Carroll, A. Delaigle, and P. Hall. Non-parametric regression estimation from data contaminated by a mixture of berkson and classical errors. Journal of the Royal Statistical Society Series B-Statistical Methodology, 69:859–878, 2007a.

D.K. McGraw and J.F. Wagner. Elliptically symmetric distributions. IEEE Trans. Inform. Theory, 14:110 – 120, 1968.

N.L. Johnson and S. Kotz. Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York, 1972.

M.E. Johnson. Multivariate Statistical Simulation. Wiley, New York, 1987.

D.N. Anderson. A multivariate linnik distribution. Statist. Proba. Lett., 14:333 – 336, 1992.

Jacek Osiewalski and Mark F. J. Steel. Robust bayesian inference in lq-spherical models. Biometrika, 80(2):456–460, 1993.

Albert W. Marshall and Ingram Olkin. Maximum likelihood characterizations of distributions. Statistica Sinica, 3(1):157–171, 1993.

Samuel Kotz, Tomaz J. Kozubowski, and Krzysztof Podgorski. The Laplace Distribution and Generalizations. Birkhäuser, 2001.

Jianqing Fan and Young K. Truong. Nonparametric Regression with Errors in Variables. The Annals of Statistics, 21(4):1900 – 1925, 1993.

Ming Chung Liu and Robert L. Taylor. A consistent nonparametric density estimator for the deconvolution problem. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 17(4):427–438, 1989.

Sam Efromovich. Density estimation for the case of supersmooth measurement error. Journal of the American Statistical Association, 92(438):526–535, 1997.

J. Berkson. Are there two regressions? Journal of the American Statistical Association, 45 (250):164–180, 1950.

R.J. Carroll, D. Rupper, C.M. Crainiceanu, and L.A. Stefanski. Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, 2 edition, 2006.

R. J. Carroll, A. Delaigle, and P. Hall. Non-parametric regression estimation from data contaminated by a mixture of berkson and classical errors. Journal of the Royal Statistical Society Series B-Statistical Methodology, 69:859–878, 2007b.

J. P. Long, N. El Karoui, and J. A. Rice. Kernel density estimation with berkson error. Canadian Journal of Statistics-Revue Canadienne De Statistique, 44(2):142–160, 2016.

Liqun Wang. Estimation of nonlinear berkson-type measurement error models. Statistica Sinica, pages 1201–1210, 2003.

Liqun Wang. Estimation of nonlinear models with berkson measurement errors. The Annals of Statistics, 32(6):2559–2579, 2004.

Liqun Wang. A unified approach to estimation of nonlinear mixed effects and berkson measurement error models. Canadian Journal of Statistics, 35(2):233–248, 2007.

Meng-Shiuh Chang and Ximing Wu. Transformation-based nonparametric estimation of multivariate densities. Journal of Multivariate Analysis, 135:71–88, 2015.